Intellectual Property Office

# Gender Profiles in Worldwide Patenting

## An analysis of female inventorship

This report was prepared by the
UK Intellectual Property Office Informatics Team
November 2016

e-mail: *informatics@ipo.gov.uk*


Intellectual Property Office
Concept House
Cardiff Road
Newport
NP10 8QQ
United Kingdom

*www.ipo.gov.uk/informatics*

**Gender profiles in worldwide patenting**

An analysis of female inventorship

November 2016

# 1 Introduction

In the UK, the Government Office for Science[1], supported by organisations such as Innovate UK[2], the Research Councils[3], the Royal Academy of Engineering[4], the IET[5] and campaigns such as WISE[6] and WiSET[7] have, for many years, been inspiring girls and women to study and build careers in the STEM fields – science, technology, engineering and mathematics. Statistical surveys from these bodies[8] highlight the impact of their work within the education sector with the number of females attaining STEM vocational qualifications increasing from 8% in 2011 to 24% in 2013. The same surveys also highlight the gender demographic transition to the workplace with women making up only 13% of the STEM workforce and women accounting for only 5.5% of engineering professionals.

Diversity statistics regarding the number of women studying STEM subjects in the education sector, up to and including degree level, are quite comprehensive because gender data is readily available regarding the number of women studying these subjects ('inputs') and those receiving qualifications ('outputs'). When looking at industry however the statistical research in this field primarily relies on 'inputs', such as the number of women employed in a given industry. Very little data is available on the 'outputs' of work undertaken by women within STEM industries but it is of great importance to governments and policymakers because of the concerns about the underrepresentation of women within science and technology.

For this reason, a recurring question that has been asked of the UK Intellectual Property Office (UK IPO) over the past few years – by UK Government colleagues, journalists and diversity and equality groups amongst others – relates to patent statistics about female inventors. Whilst absolute patent counts do not give a direct measure of innovation, they are well known as a measurable 'output' of STEM industries and it is highly desirable to analyse the inventor demographic in order to understand how inventor gender influences the patent system.

It is a legal requirement of the patent system that each inventor is named on a patent application. Disclosing further information, such as an inventor's gender, or any other protected (diversity) characteristics, is not required. This means that it has not been possible to provide statistical information about the gender of inventors named on patent applications. Until now this data has been difficult to obtain on a macroscopic level but recent name-gender inference work by several academic researchers has changed this.

---

[1] GO-Science (Government Office for Science) - *https://www.gov.uk/government/organisations/government-office-for-science*

[2] Innovate UK - *https://www.gov.uk/government/organisations/innovate-uk*

[3] Research Councils UK - *http://www.rcuk.ac.uk/*

[4] Royal Academy of Engineering - *http://www.raeng.org.uk/*

[5] IET (The Institution of Engineering and Technology) - *http://www.theiet.org/*

[6] WISE (Women in Science, Technology and Engineering) - *https://www.wisecampaign.org.uk/*

[7] WiSET (Women in Science, Engineering and Technology) - *http://www.wiset.org.uk/*

[8] For example, WISE: UK Statistics 2014 (*https://www.wisecampaign.org.uk/resources/2015/07/wise-statistics-2014*) and IET, Women in STEM: Statistics and facts (*https://communities.theiet.org/files/7976#.VbTQ7fkbJ_8*)

In March 2016 the UK IPO published a preliminary study[9], taking some baseline name-gender datasets and fusing them with GB patent data. The study showed that there was a 16% increase in the proportion of female inventors on GB patent applications in the last 10 years. It goes on to compare the proportion of British female inventors against comparator countries, and a technology breakdown of female inventors reveals a number of traditional associations.

Following the successful trial using GB patent data, this study has now been expanded to include all published patents worldwide using the European Patent Office (EPO) Worldwide Patent Statistics database, PATSTAT[10]. This report outlines the approach undertaken by the Informatics Team at the UK IPO and provides a brief study looking at the type of patent analysis that can be undertaken in terms of the gender dimension.

[9] *https://www.gov.uk/government/publications/gender-profiles-in-uk-patenting-an-analysis-of-female-inventorship*

[10] EPO Worldwide Patent Statistical Database – 2016 Spring edition *https://www.epo.org/searching-for-patents/business/patstat.html*

# 2 Previous research

There have been previous studies looking at gender in patenting but they either have a limited scope or a specific technology area/sector of focus. For example Ding *et al*[11] looks at the gender gap in patenting within the academic life sciences, using quantitative data collected for several thousand life scientists over a 30 year period alongside qualitative data from conducting face-to-face interviews. Ding found, with the limited evidence that exists about 'academic entrepreneurship', that the gender gap is of considerable magnitude. Ding states however that younger female cohorts are embracing patenting and, similarly to their male counterparts, are viewing patents as accomplishments and as a legitimate means to disseminate research. Hunt *et al*[12] took a similar approach by using the results of the US Census 2000. All respondents who had reported having a bachelor's degree or higher were contacted to establish if they had applied for a US patent since 2008. Those who had were asked how many had been commercialised or licensed, on the author's assumption that these are important for economic growth. Although both Ding and Hunt have similar conclusions they have a relatively narrow scope because they are academic-centric and rely heavily on anecdotal information.

Various studies have tried to explain the cause of the gender gap in patenting, with Ding concluding that women have fewer contacts in industry and that female scientists suffer from an attainment gap along at least three important dimensions: productivity, recognition and reward. Some of these studies used patent data to try and explain the gender gap but most are relatively narrow in scope. For example Jung *et al*[13] covers Swedish patent data alone, Mauleón *et al*[14] covers Spanish patent data and Whittington *et al*[15] is based on data from the US Patent and Trademark Office (USPTO).

Clearly the biggest difficulty when using patent data is that there is no easy way to establish the gender of the inventors due to limitations of the data, and all previous methodologies rely on using additional databases of common names to identify gender. The study by the National Women's Business Council[16] is limited to USPTO data between 1975 and 2010 but managed to assign 94% of patents using commercially available data to match names for men and women. The study also paid particular care not to overlook the contributions of immigrant US-based women by including uncommon Chinese, Korean, Indian, Japanese and European names. The remaining 6% were then allocated gender based on ratio of patents granted to males and females on a year-by-year basis.

Sugimoto *et al*[17] conducted a similar study using 4.6 million USPTO patents issued between 1976 and 2013. Inventors from 185 countries applied to the USPTO during this

---

[11] Ding, W. *et al*, 2006. Gender Differences in Patenting in the Academic Life Sciences. Science, 313 (5787), 665-687

[12] Hunt, J. *et al*, 2012. Why don't women patent?, NBER Working Paper No. 17888

[13] Jung, T. *et al*, 2012, Demographic patterns and trends in patenting: Gender, age, and education of inventors. Centre for Innovation, Research and Competence in the Learning Economy (CIRCLE), Lund University, Paper no. 2012/05

[14] Mauleón, E. *et al*, 2009. Male and female involvement in patenting activity in Spain. Scientometrics (2010) 83:605–621

[15] Whittington, K. *et al*, 2008, Women inventors in context: Disparities in Patenting across Academia and Industry. Gender & Society, Vol. 22 No. 2, (2008) 194-218

[16] National Women's Business Council, 2012. Intellectual Property and Women Entrepreneurs. *https://www.nwbc.gov/sites/default/files/IP%20&%20Women%20Entrepreneurs.pdf*

[17] Sugimoto, C. *et al*, 2015, The Academic Advantage: Gender Disparities in Patenting. PLoS ONE 10(5): e0128000.

period and the accuracy of the gender assignation table varies by inventor country, with a high of 90% for the US and a low of 66% for India. Across all USPTO patents used, 86% of inventor-patent combinations were assigned to a gender. Sugimoto had a few key findings, such as “patents to which women – and in particular academic women – contributed are associated with a higher number of International Patent Classification (IPC) codes and co-inventors than men, suggesting higher inter-disciplinarity in female patenting”. It is however difficult to draw any firm conclusions from analysing the number of IPC codes as classifications are determined by the patent examiner and are therefore subjective, causing disparities in the allocation of patent classifications between and within patent offices[18]. Sugimoto also found that male dominance in patenting is found in nearly every country and, in every technological area, female patenting is proportionally more likely to occur in academic institutions than in corporate or government environments.

Frietsch *et al*[19] analysed patent applications rather than granted patents claiming applications “are published earlier and better reflect technological competiveness of an invention”. Frietsch focused on applications filed at the European Patent Office (EPO) over five priority years (1993, 1996, 1998, 2000 and 2001) and excluded countries where successful name matching was less than 85%. The dataset was therefore narrowed to 14 countries who apply at the EPO out of a possible 25. This study also looks at non-patent academic/scientific literature which shows one clear finding, namely that the proportion of women decreases with an increasing level of education and seniority. It states “in 2003 women completed 43% of the doctoral theses that were examined and 32% of the habilitations. Yet there were only 15% female professors employed at universities”.

The study by Naldi *et al*[20] managed to identify the gender of authors of publications or inventors of a patent in more than 90% of cases by developing their own proprietary First Name Data Base (FNDB). Naldi used EPO publications from 1998 relating to 100,000 inventors for the six EU countries selected (UK, France, Germany, Italy, Spain and Sweden), as well as 30,000 authors of scientific publications. Naldi breaks down patents by gender as well as analysing the gender split by industry sector.

In 2012 the OECD created a database[21] giving top level statistics on the gender of inventors limited to top level figures for OECD Member countries plus several other selected countries.

In March 2016 the UK IPO published a preliminary study[22], taking some baseline name-gender datasets and fusing them with GB patent applications filed between 1978 and 2015. The UK IPO report, and all the others mentioned above, are limited in their scope, whether in terms of patent jurisdiction coverage or date range limitations. This previous UK IPO report suggested that only when the work is expanded to include worldwide patent data will it be possible to accurately compare the gender disparities worldwide and draw further conclusions about women in patenting. This defined the scope of the study discussed in this report which aimed to fill this gap in the literature.

---

[18] As explained in more detail in The Patent Guide, published by the UK IPO in 2015. *https://www.gov.uk/government/publications/the-patent-guide*

[19] Frietsch, R. *et al*, 2009, Gender-specific patterns in patenting and publishing. Research Policy 38 (2009) 590-599

[20] Naldi, F. *et al*., 2002. Scientific and Technological Performance by Gender, Vol 1 and Vol 2, European Commission

[21] *http://www.oecd.org/gender/data/dowomenpatent.htm*

[22] *https://www.gov.uk/government/publications/gender-profiles-in-uk-patenting-an-analysis-of-female-inventorship*

# 3 Data sources and methodology

## 3.1 Patent data

The preliminary study undertaken by the UK IPO[9] focused on fusing baseline name-gender datasets with GB patent data held on the UK patents register. To expand this work to include all patents worldwide required the use of a different source of patent data. The UK IPO is keen to make the outputs of this study available for others to use and, since the European Patent Office (EPO) Worldwide Patent Statistical Database (PATSTAT)[10] is widely used by patent researchers and academics worldwide, it was the natural choice for this study. Containing bibliographic data relating to more than 90 million patent documents from leading industrialised and developing countries[23], this database provides a comprehensive collection of data which can be set up and queried using Structured Query Language (SQL).

## 3.2 Name-gender data

A small number of recent academic research projects looking at inferring gender from name data have made it possible to infer inventor gender on patent applications. This study primarily focuses on two different methodologies, one originating from Massachusetts Institute of Technology (MIT)[24] (hereinafter known as the Matias methodology) and the other from Peking University/NYU Polytechnic School of Engineering/Max Planck Institute for Software Systems[25] (hereinafter known as the Tang methodology).

The Matias methodology originated from research at MIT undertaken in collaboration with Bocoup and funded by the Knight Foundation.  It involved collecting open source annual birth data from the US Social Security Administration and the UK Office for National Statistics (ONS) into a single database. US data from the US Social Security Administration provides records for name and gender by year for births between 1880 and 2011. UK ONS data records births for England and Wales between 1996 and 2011, with Scotland (2009 and 2010 only) and Northern Ireland (1997-2011) recorded separately. The resulting US and UK name lists each comprise the number of male and female entries and the number of years in which each name appears. For the purposes of this study the UK IPO combined both the US and UK name lists for further analysis.

The Tang methodology comprises collaborative research by Peking University, NYU Polytechnic School of Engineering and the Max Planck Institute for Software Systems. It involved crawling Facebook® public profile pages for millions of users to generate an annotated name-gender list. The research goes on to use this name-gender list to infer gender information for users who do not explicitly specify their gender and then provides some analysis of gender characteristics and gender behaviour in Facebook®. For the

---

[23] https://www.epo.org/searching-for-patents/business/patstat.html

[24] https://github.com/OpenGenderTracking/globalnamedata - MIT PhD research (Matias, N.) undertaken in collaboration with Bocoup and funded by the Knight Foundation

[25] Tang, C. et al (2011) What's in a Name: A Study of Names, Gender Inference, and Gender Behavior in Facebook. *Database Systems for Advanced Applications*. 6637. p. 344-356

purposes of this study the UK IPO was only interested in using the annotated name-gender list that was populated using web-crawling to extract the user-disclosed name and gender data from these Facebook® public profiles. This name-gender list comprises the number of sampled Facebook® users having each name (all one-letter names, names without a vowel, and names referenced only once were removed), the number of times it is labelled as male and the number of times it is labelled as female.

Both the Matias and Tang methodologies provide open source datasets listing names alongside a count of how many entries are male and female, as shown in Table 1. The UK IPO used the names listed in these two open source databases to infer a gender with an assigned confidence score based on the number of male/female entries compared to the total number of entries. For example, in Table 1 the gender of Samantha would be inferred to be female with a 100% confidence score and Matty would be inferred to be male with a 94% confidence score.

**Table 1: Example of name list entry format**

| Name | Entries | Male | Female |
|---|---|---|---|
| Samantha | 11906 | 0 | 11906 |
| Matty | 116 | 109 | 7 |

In order to provide a high-quality dataset for analysis, the confidence threshold for gender inference was set quite high at 95%. A combination name-gender dataset consisting of data from both the Matias and Tang methodologies was used for this study alongside additional manual quality check data cleaning undertaken by the UK IPO using a variety of additional online data verification sources[26]. The resulting name-gender dataset contains 102,777 unique first names with an inferred gender.

[26] Including, but not limited to, http://www.behindthename.com/, http://www.gender-api.com/ and the data sources listed in Appendix B of Naldi (2002) Vol. 2 and Table S1 of the Supplementary Information accompanying Sugimoto (2013) (http://www.nature.com/news/bibliometrics-global-gender-disparities-in-science-1.14321)

## 3.3 Methodology

### 3.3.1 Data extraction

Bibliographic information relating to the patent applications contained in PATSTAT[10] are stored in different relational tables comprising information about the patent publications, applicants, inventors, classifications, priorities, families and citations, amongst others. All of this information can be linked together using either a unique application identifier or person identifier[27]. Sequence numbering is used to separate person identifiers relating to applicants and inventors[28] and thus it is possible to extract a list of all person identifiers with sequence numbering that shows that they are inventors and not applicants (*e.g.* companies).

PATSTAT contains 38,408,782 unique inventor person identifiers. These unique inventor identifiers and the associated person name formed the PATSTAT data to be extracted using SQL for this study.

### 3.3.2 Data cleaning and data linking

Inventor information used for the preliminary study of GB patent data is held on the UK patents register. This inventor data is relatively clean because it follows a standard name format (*i.e.* surname, given name). Unfortunately the inventor name data held in PATSTAT is not so clean because the person data offered to the EPO from the individual patent registers of national and regional patent offices around the world is not in a standard format. For example, inventor name data for Swedish and Swiss national patents appears to be in the 'given name(s) surname' format whereas inventor name data for French and German national patents appears to be in the 'surname given name(s)' format. The inventor name data is even more complicated because the name format can change over time; for example, the person data on GB patents appears to be in the 'given name(s) surname' format over certain time periods in history but in the 'surname given name(s)' format in others. Inventor name samples were taken for each jurisdiction and name formats were manually checked on these samples with a look-up table created.

There are also some other issues with the inventor information contained within PATSTAT. There are 85,790,188 patent applications in PATSTAT but 26,679,490 have no inventor information at all, so inventor information only actually exists for 68.9% of all PATSTAT patents. As mentioned above, PATSTAT contains 38,408,782 unique inventor person identifiers, but 1,610,134 of these do not contain full first name information (*e.g.* inventor initial(s) only[29]). This leaves 36,798,648 unique inventor person identifiers with suitable name information from which attempt to infer inventor gender.

Although PATSTAT contains unique inventor identifiers, there may be multiple inventor names associated with one inventor identifier because of the name structure by application authority (jurisdiction) issue mentioned above. For example, the inventor name for inventor identifier *12345* may appear as *Joe Bloggs* or *Bloggs Joe* depending on the application authority. To address this issue, every inventor name format was extracted for each unique inventor identifier; the number of applications for each name format for each

---

[27] In PATSTAT the applicant identifier is *appln_id* and the person identifier is *person_id*

[28] In PATSTAT the inventor sequence number (*invt_seq_nr*) is 0 for applicants and 1 for inventors

[29] For this reason, the decision was taken to only attempt to infer gender from first names consisting of three or more characters

inventor identifier were counted and the assumption was then made that the most regularly occurring name format is most likely to be the one to use to extract the first name and match to the name-gender dataset. There are however a small number of instances where the number of applications may be the same for each name format; for example, *Joe Bloggs* may have 4 CN, 5 US and 3 DE patent applications and *Bloggs Joe* may have 12 GB patent applications. In this instance it is unclear if *Joe* or *Bloggs* is the first name that is trying to be matched to the name-gender dataset. To address this, both names are matched to the name-gender dataset; in this instance the name *Joe* is inferred to be male (M) and *Bloggs* is unassigned (U) since it does not appear in the name-gender dataset. Any M-M or M-U combination can be correctly inferred to be male and any F-F or F-U combination inferred to be female. The one exception is the small number of instances where there is an M-F combination (*e.g.* the name *Martha James*) in which case both inferred genders are ignored and the unique inventor identifier is assigned an unknown (null) gender.

To assess the quality of the PATSTAT inventor name data, the inventor names associated with each of the 36,798,648 unique inventor identifiers in PATSTAT capable of having inventor gender inferred were passed through the same SQL first name extraction stored procedure and matched to the same name-gender assignment dataset that were used for the preliminary study using GB patents. This revealed a number of challenges with PATSTAT worldwide inventor name data that were not encountered previously when using GB data only, including:

- Name suffixes (often found in US inventor names), *e.g.* II, III, IV, Sr, Jr.
- Honorifics, *e.g.* Dr, Prof.
- Qualifications appearing in the inventor name field, *e.g.* Dipl-Ing is common for German inventors.
- “Deceased” appearing at the end of the inventor name field.
- Poor coverage for East Germany (DD) and Soviet Union (SU) patents was caused by the respective jurisdiction code appearing at the end of the inventor name field.

All of these issues resulted in various incorrect first names being extracted, which ultimately resulted in a poor match rate with the name-gender dataset. The name cleaning algorithm was then updated to take account of these issues and improve the quality of first name extraction from the inventor name field.

Iterative name cleaning improvements[30] to tens of millions of inventor names proved time-intensive so the decision was taken to use a combination of techniques to implement a data cleaning algorithm using both SQL querying and Python scripting.

Further data cleaning was undertaken to handle various exceptions, including the use of country-specific dictionaries for known unisex names. For example, *Jean* is generally considered to be a female name except in French-speaking countries where *Jean* is male. In this instance, *Jean* is inferred to be male if the reported inventor country is France, Belgium, Canada etc, but otherwise is inferred to be female if there is any other non-

[30] Various alternative data sources including, but not limited to, *http://www.behindthename.com/*, *http://www.gender-api.com/* and the data sources listed in Table S1 of the Supplementary Information accompanying Sugimoto (2013) (*http://www.nature.com/news/bibliometrics-global-gender-disparities-in-science-1.14321*) were used to address holes in the coverage of the dataset; for example, manual name-gender checking for the top unmatched names and the top unmatched names by jurisdiction and inventor country

French speaking inventor country information. Any instances of *Jean* as the inventor name but with no accompanying inventor country data is assigned an unknown (null) gender. Similar issues occur with other unisex names such as *Andrea* and *Nicola* which are generally female except in Italy where they are male names, as well as *Patrice*, *Simone*, *Marian* and *Michele*, amongst others.

#### 3.3.3 Research output

Accompanying this report, the UK IPO has published[31] a table which can be used in conjunction with PATSTAT containing the inventor gender disaggregation for 26,997,717 identifiers in PATSTAT.

[31] *https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship*

# 4 Analysis of female inventors

## 4.1 Dataset overview

An overview of the final dataset used for this study is provided in Figure 1. It shows that 27.0m of the 36.8m ‘matchable’ unique inventor identifiers (73.4%) in PATSTAT (*i.e.* inventor first names with initials only removed) have successfully had a gender inferred with a 95% certainty. This equates to 70.3% of all unique inventor identifiers in PATSTAT (27.0m out of 38.4m unique inventor identifiers). The summary tables that follow (Table 2 to Table 4) show the top male and female inventor names in PATSTAT, by the number of unique inventor identifiers, total patent applications and inventor country respectively.

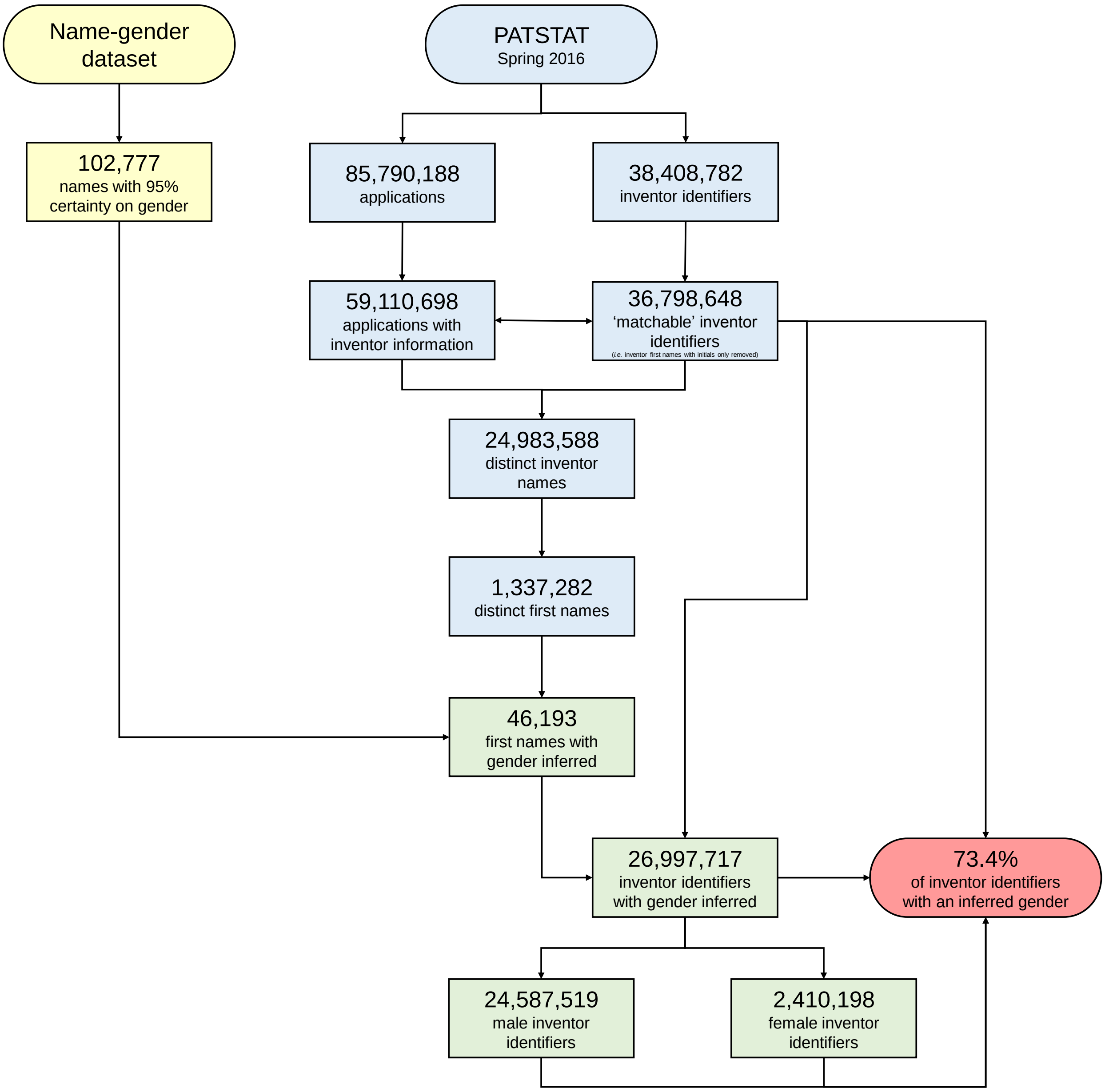



Figure 1: Dataset overview

### Table 2: Names of the top 20 male and female unique inventor identifiers

| Rank | Male names | Unique inventor identifiers |
|---|---|---|
| 1 | John | 475,112 |
| 2 | Robert | 402,528 |
| 3 | Michael | 390,270 |
| 4 | David | 383,531 |
| 5 | James | 306,289 |
| 6 | William | 287,773 |
| 7 | Thomas | 278,326 |
| 8 | Peter | 266,223 |
| 9 | Richard | 242,181 |
| 10 | Paul | 205,966 |
| 11 | Charles | 161,881 |
| 12 | Joseph | 159,530 |
| 13 | Mark | 158,426 |
| 14 | Daniel | 142,641 |
| 15 | Vladimir | 136,562 |
| 16 | George | 129,961 |
| 17 | Martin | 116,039 |
| 18 | Frank | 112,370 |
| 19 | Stephen | 106,879 |
| 20 | Hans | 106,705 |

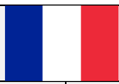
| Rank | Female names | Unique inventor identifiers |
|---|---|---|
| 1 | Maria | 31,583 |
| 2 | Elena | 20,837 |
| 3 | Anna | 19,838 |
| 4 | Susan | 18,598 |
| 5 | Mary | 17,945 |
| 6 | Galina | 15,839 |
| 7 | Irina | 15,629 |
| 8 | Barbara | 14,558 |
| 9 | Tatyana | 14,364 |
| 10 | Jennifer | 14,353 |
| 11 | Anne | 14,192 |
| 12 | Christine | 13,876 |
| 13 | Lyudmila | 13,629 |
| 14 | Elizabeth | 13,556 |
| 15 | Valentina | 13,292 |
| 16 | Karen | 13,226 |
| 17 | Olga | 12,531 |
| 18 | Nina | 12,383 |
| 19 | Catherine | 11,820 |
| 20 | Patricia | 11,528 |

### Table 3: Top 20 male and female inventor names

| Rank | Male names | Patent applications |
|---|---|---|
| 1 | John | 1,165,493 |
| 2 | Michael | 1,057,292 |
| 3 | Robert | 1,011,510 |
| 4 | David | 964,706 |
| 5 | Hiroshi | 789,375 |
| 6 | Thomas | 781,780 |
| 7 | James | 742,271 |
| 8 | Peter | 718,643 |
| 9 | William | 702,571 |
| 10 | Takashi | 610,972 |
| 11 | Richard | 589,473 |
| 12 | Paul | 533,130 |
| 13 | Vladimir | 446,323 |
| 14 | Takeshi | 439,521 |
| 15 | Mark | 415,681 |
| 16 | Hiroyuki | 403,736 |
| 17 | Charles | 385,771 |
| 18 | Joseph | 383,977 |
| 19 | Kenji | 379,440 |
| 20 | Daniel | 362,507 |

| Rank | Female names | Patent applications |
|---|---|---|
| 1 | Maria | 67,937 |
| 2 | Elena | 43,542 |
| 3 | Anna | 42,002 |
| 4 | Susan | 40,516 |
| 5 | Mary | 37,564 |
| 6 | Barbara | 33,751 |
| 7 | Christine | 32,684 |
| 8 | Anne | 32,240 |
| 9 | Jennifer | 31,820 |
| 10 | Irina | 31,367 |
| 11 | Galina | 30,703 |
| 12 | Karen | 28,648 |
| 13 | Catherine | 28,419 |
| 14 | Elizabeth | 27,847 |
| 15 | Lyudmila | 26,875 |
| 16 | Valentina | 26,686 |
| 17 | Tatyana | 26,547 |
| 18 | Tomoko | 26,078 |
| 19 | Olga | 25,305 |
| 20 | Nina | 25,145 |

### Table 4: Top 20 male and female names by inventor country

|  | |  | |  | |
|---|---|---|---|---|---|
| ♂ | ♀ | ♂ | ♀ | ♂ | ♀ |
| David | Susan | Philippe | Isabelle | Thomas | Sabine |
| John | Helen | Jean | Catherine | Michael | Ulrike |
| Michael | Sarah | Michel | Nathalie | Peter | Claudia |
| Peter | Elizabeth | Pierre | Anne | Wolfgang | Andrea |
| Andrew | Alison | Alain | Christine | Andreas | Susanne |
| Paul | Catherine | Bernard | Sylvie | Klaus | Petra |
| Robert | Jane | Jacques | Sophie | Martin | Heike |
| Richard | Karen | Patrick | Marie | Stefan | Birgit |
| Stephen | Julie | Laurent | Francoise | Gerhard | Karin |
| James | Mary | Christian | Veronique | Bernd | Anja |
| Christopher | Anne | Eric | Florence | Werner | Barbara |
| William | Caroline | Jean-Pierre | Valerie | Hans | Monika |
| Ian | Margaret | Olivier | Sandrine | Manfred | Ursula |
| Mark | Emma | Christophe | Claire | Frank | Christine |
| Alan | Claire | Claude | Pascale | Christian | Elke |
| Martin | Ashley | Daniel | Caroline | Dieter | Maria |
| Anthony | Louise | Pascal | Corinne | Juergen | Martina |
| Simon | Gillian | Thierry | Cecile | Ulrich | Kerstin |
| Philip | Sandra | Francois | Martine | Helmut | Gisela |
| Brian | Rachel | Gerard | Elisabeth | Matthias | Brigitte |

|  | |  | |  | |
|---|---|---|---|---|---|
| ♂ | ♀ | ♂ | ♀ | ♂ | ♀ |
| John | Susan | Hiroshi | Tomoko | Vladimir | Elena |
| David | Mary | Takashi | Akiko | Aleksandr | Irina |
| Robert | Jennifer | Hiroyuki | Yuko | Oleg | Galina |
| Michael | Karen | Takeshi | Keiko | Sergej | Svetlana |
| James | Elizabeth | Kenji | Noriko | Viktor | Julija |
| William | Lisa | Satoshi | Hiroko | Nikolaj | Ljudmila |
| Richard | Linda | Makoto | Tomomi | Mikhail | Marina |
| Thomas | Barbara | Masahiro | Naoko | Jurij | Anna |
| Mark | Patricia | Koji | Junko | Aleksej | Valentina |
| Paul | Nancy | Atsushi | Kyoko | Andrej | Marija |
| Joseph | Laura | Kenichi | Yukiko | Evgenij | Nadezhda |
| Charles | Deborah | Masayuki | Kaori | Igor | Nina |
| Daniel | Christine | Osamu | Yumiko | Anatolij | Ekaterina |
| Steven | Maria | Kazuo | Miyuki | Valerij | Larisa |
| Jeffrey | Amy | Koichi | Naomi | Dmitrij | Tatyana |
| Peter | Lynn | Shinichi | Masako | Boris | Eva |
| Stephen | Margaret | Yuji | Yoshiko | Tat | Olga |
| Christopher | Jean | Toshio | Ayako | Gennadij | Tamara |
| Brian | Ann | Hiroaki | Mayumi | Yurij | Vera |
| George | Kathleen | Yasuhiro | Sachiko | Ivan | Ljubov |

## 4.2 Historical profiling

Figure 2 shows the annual percentage of female inventors on all published patent applications between 1915 and 2015. There is a clear increase in the proportion of female inventors from around 2-3% during most of the first half of the 20th Century to around 6-11% since 2000. Although absolute numbers remain relatively low, the last 15 years has seen the proportion of female inventors worldwide increase over 60% (7.1% in 2001; 11.5% in 2015).

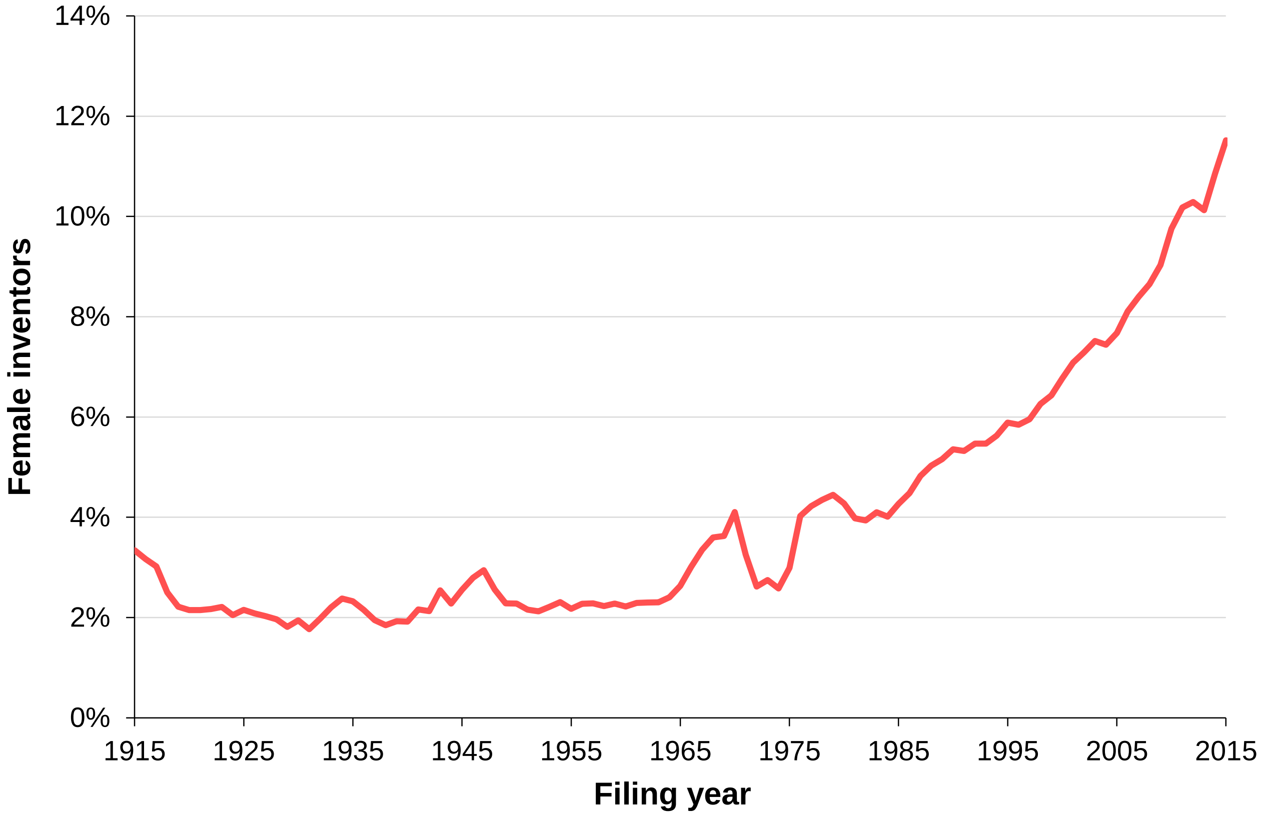



**Figure 2: Female inventors, 1915-2015**

Three indicators are commonly used, namely female 'participation'[32], female 'contribution'[33] and 'total count'[34] of female inventors. The data presented in Figure 2, and all other analysis presented in section 4, is based on 'total count' data[35]. When female contribution is analysed, the results over the last 40 years display the same trend as shown in Figure 2 but the female contribution using fractional counting is approximately 0.5%–1% lower each year than the total count data (*e.g.* female inventors in 2010 represent 9.8% using female

32 Number of patents with at least one female inventor

33 Fractional counting measuring the female involvement to the production of a patent assuming that each inventor concurred with the same effort, *e.g.* a patent with five inventors of which two are female will have a female contribution of 2/5

34 Total number of female inventors

35 Inventor gender ratios using total count data is presented in section 4. Full fractional count data to analyse female contribution, and also female participation, can be downloaded from https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship

total count data and 8.9% using female contribution fractional counting).

Figure 3 is a subset of the data presented in Figure 2 and shows the annual percentage of female inventors on published patent applications over the past 25 years (1990-2015) by six select inventor countries[36]. There are noticeable differences in the percentage of female inventors between countries. For example, since 1990 the percentage of female inventors in France is generally at least 3% higher than it is for female inventors in the UK.

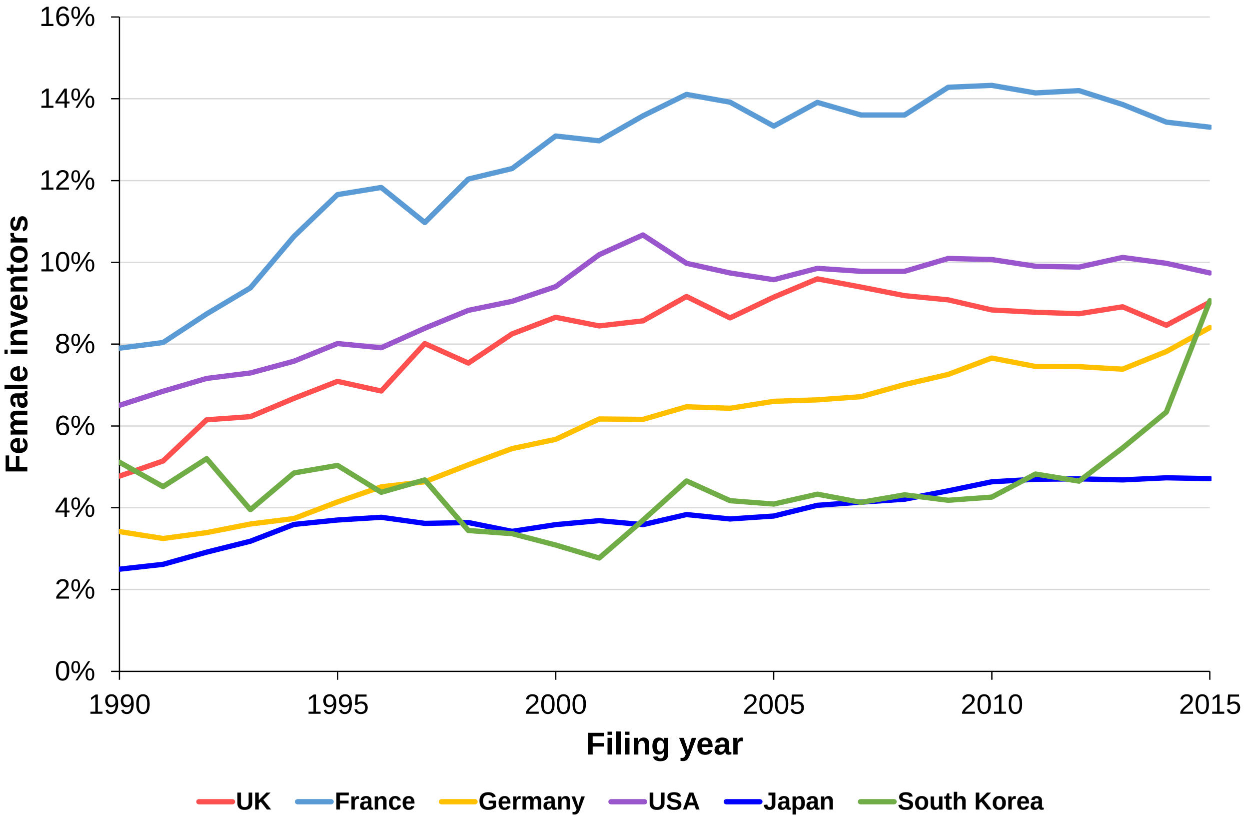



**Figure 3: Female inventors by select inventor country, 1990-2015**

[36] When applying for a patent you do not need normally to provide any nationality information so any references to inventor country are simply the country of residence of the inventor. This is not perfect given the mobility of inventors however it is a good first estimate

## 4.3 Geoanalytics

The differences shown previously in Figure 3 highlight clear disparity between female inventors in different inventor countries. Table 5 and Table 6 investigate this in more detail, respectively showing the inventor gender ratios[37] for the top patent jurisdictions[38] and top inventor countries[39] worldwide.

**Table 5: Top jurisdictions**

| Jurisdiction | | Patent applications | Inventor identifiers | Inventor gender ratio | | Proportion of dataset | Gender known |
|---|---|---|---|---|---|---|---|
| | | | | Male | Female | | |
| US | United States of America | 25,080,307 | 20,601,443 | 93.16% | 6.84% | 18.92% | 82.11% |
| JP | Japan | 24,679,416 | 20,925,258 | 96.85% | 3.15% | 18.62% | 84.78% |
| CN | China | 24,674,386 | 9,428,007 | 88.53% | 11.47% | 18.62% | 38.19% |
| DE | Germany | 9,316,566 | 7,834,315 | 95.79% | 4.21% | 7.03% | 83.99% |
| EP | European Patent Office | 8,777,596 | 7,598,235 | 92.21% | 7.79% | 6.62% | 86.56% |
| KR | South Korea | 5,996,948 | 2,793,060 | 94.70% | 5.30% | 4.52% | 46.57% |
| SU | Soviet Union (USSR) | 3,693,181 | 3,041,148 | 89.27% | 10.73% | 2.79% | 82.34% |
| AU | Australia | 3,550,095 | 2,301,396 | 90.60% | 9.40% | 2.68% | 64.60% |
| CA | Canada | 3,525,615 | 3,108,855 | 91.95% | 8.05% | 2.66% | 88.08% |
| TW | Taiwan | 2,212,149 | 999,233 | 92.44% | 7.56% | 1.67% | 45.17% |
| FR | France | 1,942,911 | 1,560,476 | 93.75% | 6.25% | 1.47% | 78.50% |
| GB | United Kingdom | 1,805,127 | 1,563,284 | 95.96% | 4.04% | 1.36% | 86.49% |
| RU | Russian Federation | 1,694,191 | 948,789 | 84.98% | 15.02% | 1.28% | 56.00% |
| AT | Austria | 1,652,202 | 1,365,601 | 92.78% | 7.22% | 1.25% | 82.53% |

Table 5 shows that Japanese and GB patents have the lowest proportion of female inventors (<4%) of the top 14 patenting jurisdictions in the world (the worldwide average across all jurisdictions is 6.9%); with gender inference rates of over 80% in each case, it can be concluded that this is a fairly accurate representation of Japanese and GB patents as a whole. It is however important not to draw too many conclusions from jurisdiction information because anyone in the world can file for a patent in any jurisdiction whereas studying the self-reported inventor residency (Table 6) provides a more meaningful way to analyse the proportion of female inventors in different countries.

[37] Inventor gender ratios using total count data is presented here. Full fractional count data to analyse female contribution, and also female participation, can be downloaded at *https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship*

[38] Top jurisdictions selected as those that account for >1% of all patents in PATSTAT

[39] Top inventor countries selected as those that account for >0.5% of all patents in PATSTAT

**Table 6: Top inventor countries**

| Inventor Country | | Patent applications | Inventor identifiers | Inventor gender ratio | | Proportion of dataset | Gender known |
|---|---|---|---|---|---|---|---|
| | | | | Male | Female | | |
| US | United States of America | 19,399,694 | 16,443,297 | 91.28% | 8.72% | 14.64% | 84.76% |
| JP | Japan | 9,795,869 | 8,542,067 | 96.27% | 3.73% | 7.39% | 87.20% |
| DE | Germany | 8,568,615 | 7,650,425 | 94.48% | 5.52% | 6.46% | 89.28% |
| KR | South Korea | 5,247,715 | 1,526,520 | 95.56% | 4.44% | 3.96% | 29.09% |
| CN | China | 4,341,568 | 1,211,085 | 89.94% | 10.06% | 3.28% | 27.90% |
| FR | France | 2,590,087 | 2,276,998 | 88.29% | 11.71% | 1.95% | 87.84% |
| GB | United Kingdom | 2,151,629 | 1,924,717 | 92.74% | 7.26% | 1.62% | 89.45% |
| SU | Soviet Union (USSR) | 1,862,730 | 1,631,818 | 88.91% | 11.09% | 1.41% | 87.60% |
| TW | Taiwan | 1,663,221 | 193,233 | 90.19% | 9.81% | 1.25% | 11.62% |
| RU | Russian Federation | 1,191,600 | 918,380 | 84.31% | 15.69% | 0.90% | 77.07% |
| CH | Switzerland | 1,107,248 | 840,751 | 93.26% | 6.74% | 0.84% | 75.93% |
| CA | Canada | 1,054,656 | 893,528 | 91.34% | 8.66% | 0.80% | 84.72% |
| IT | Italy | 1,048,578 | 902,402 | 88.37% | 11.63% | 0.79% | 86.06% |
| NL | Netherlands | 903,190 | 687,057 | 92.81% | 7.19% | 0.68% | 76.07% |
| SE | Sweden | 734,895 | 573,107 | 90.95% | 9.05% | 0.55% | 77.98% |
| *Unknown* | | *64,748,228* | *43,390,638* | *94.20%* | *5.80%* | *48.85%* | *66.85%* |

Table 6 shows the top 15 inventor countries using the self-reported[40] inventor residency information. There are some clear data coverage limitations with this because no inventor country data exists for almost half of the patents in PATSTAT. Although no cross-validation can be performed, it is still possible to draw comparisons from the data that does exist because the inventor gender ratio for each country is likely to stay roughly the same.

The data in Table 6 indicates that France, China and Russia (including the USSR) have the highest proportion of female inventors of the top inventor countries, with Japan, Germany and South Korea having the lowest (the worldwide average is 7.2%). Table 6 also highlights one of the limitations of this study because, as with the previous research in this field (see section 2), the accuracy of patent analysis is only as good as the quality of name-gender assignments in the underlying master dataset that the patent data is matched to. It is clear from Table 6 that gender inference rates are low for Chinese, Taiwanese and Korean inventor names; for example, only 27.9% of all patents from Chinese inventors have successfully had a gender inferred with a 95% certainty. This is primarily due to handling non-Latin alphabet names and machine-transliterations of non-Latin characters and further improvement to country-specific dictionaries is required (explained in more detail in section 5).

Figure 4 is a choropleth map showing the proportion of female inventors by inventor country[41]. Several countries in South America and Africa have a relatively high percentage of female inventors but the absolute number of inventors in these countries is low. Figure 5

[40] The limitations of which are explained in more detail in The Patent Guide, published by the UK IPO in 2015. https://www.gov.uk/government/publications/the-patent-guide
[41] The darker the colour, the higher the percentage of female inventors

shows the same data but is limited to the top 50 inventor countries (>22,000 inventors), making it easier to distinguish the impact of female inventorship across the main patenting countries.

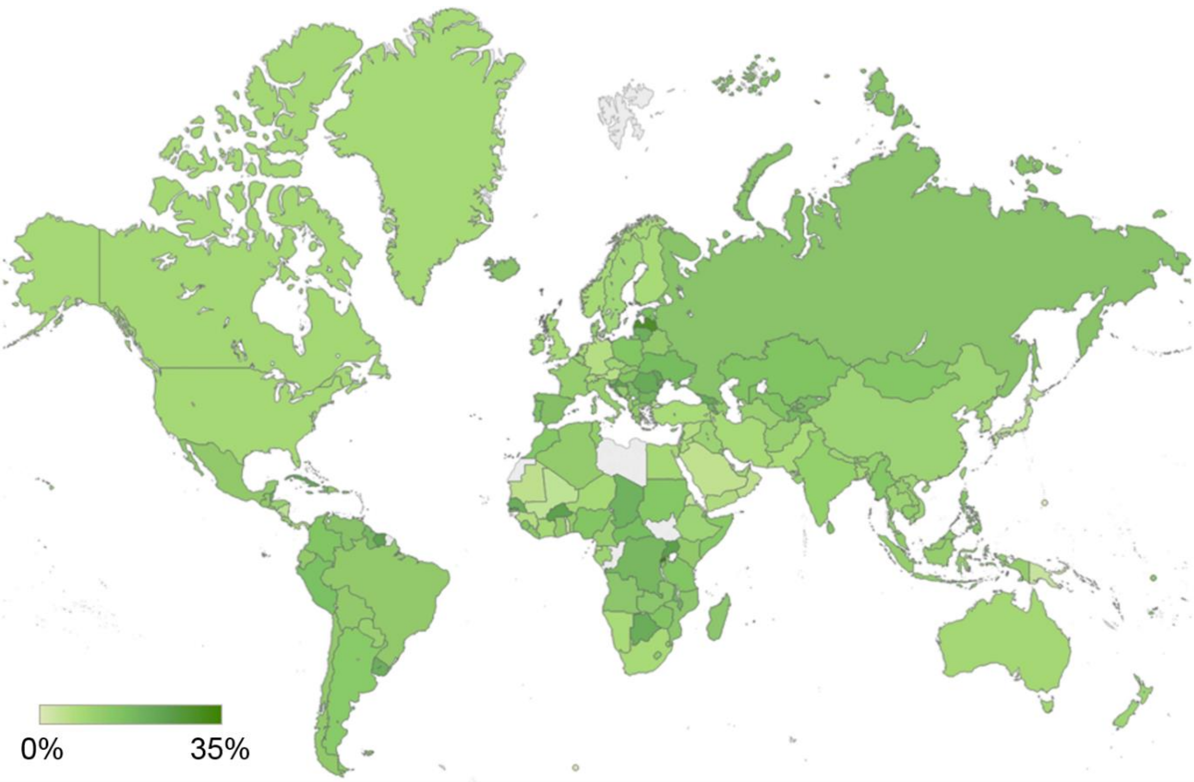



**Figure 4: Choropleth map showing proportion of female inventors by inventor country**

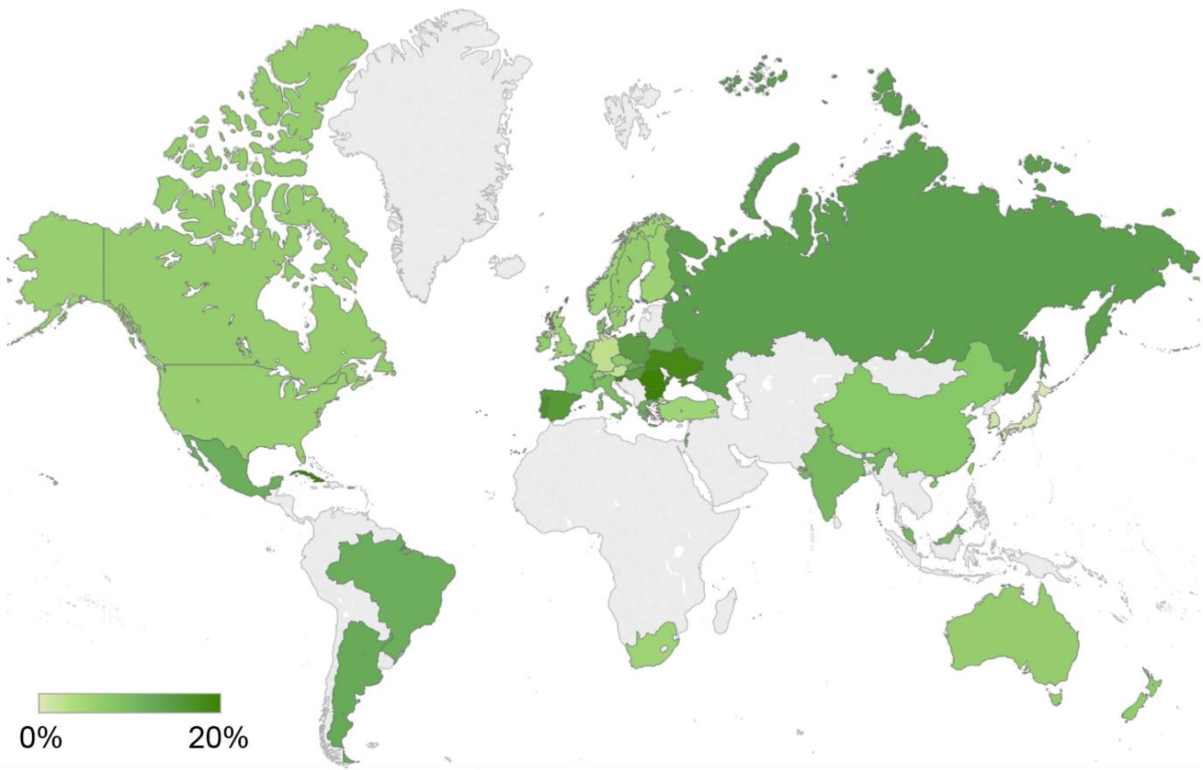



**Figure 5: Choropleth map showing proportion of female inventors by top 50 inventor countries**

## 4.4 Teamworking

One area of particular interest when looking at inventor gender is in the analysis of the number of inventors on each patent application. Each patent application can range from having one named inventor (a lone/individual inventor) to multiple inventors (working collaboratively as part of a team). By linking the inferred gender of each named inventor and the number of inventors listed on each patent application, analysis could then be undertaken on whether female inventors are more likely to work on their own, as part of an all-female team, or as part of a mixed team.

In the dataset analysed, 87.0% of patents involve only men (48.5% individual males, 38.5% all-male teams), 2.2% involve only women (2.0% individual females, 0.2% all-female teams) and the remaining 10.8% involved mixed teams. As shown previously, there has been big increases in the amount of female patenting in recent years and the results in Figure 6 show the percentage of female inventors on patents filed between 1975 and 2015 split by inventor type.

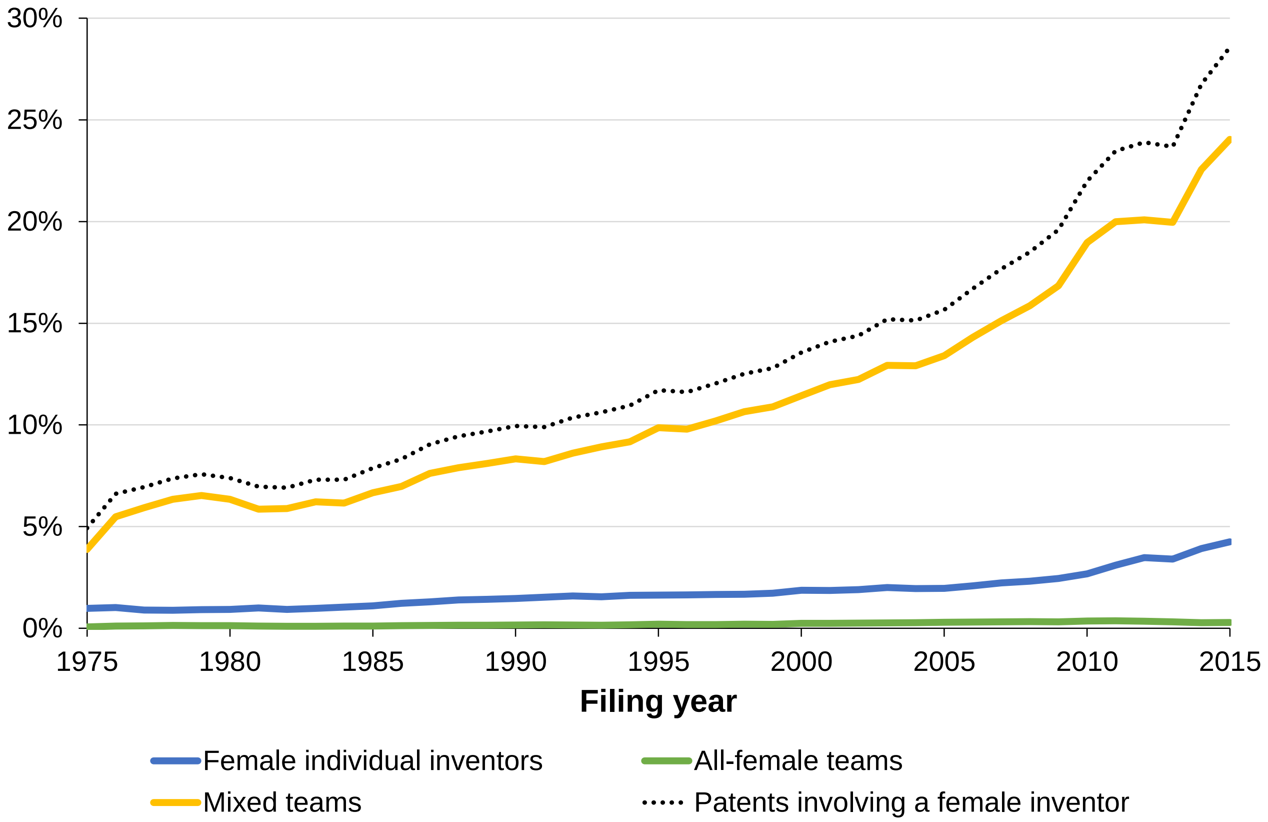



**Figure 6: Female inventors by inventor type**

The overall proportion of patents involving a female inventor has increased more than 500% from 4.9% in 1975 to over 28% in 2015. This is predominantly due to year-on-year increases in the number of female inventors working as part of mixed teams. Individual female inventors accounted for less than 1% of patents between 1975 and 1985 but this has slowly increased to average 3% since 2005. The number of all-female teams has increased over four times since 1975 but the absolute numbers are very low with 0.06% of patents coming from all-female teams in 1975 and only 0.28% in 2015.

The pie charts in Figure 7 show a clear gender disparity and compares the split of inventor types on PATSTAT patents filed in 1984 and 2014. Although the positive trends shown in Figure 6 appear promising, the increasing proportion of female inventors is slow, and the absolute numbers are still very low. Figure 7 shows the bigger picture and is a more accurate reflection of the gender disparity within patenting worldwide; 73% of all patent applications worldwide in 2014 are still from all-male inventors and this rises to almost 96% when mixed teams (with at least one male inventor) are considered.

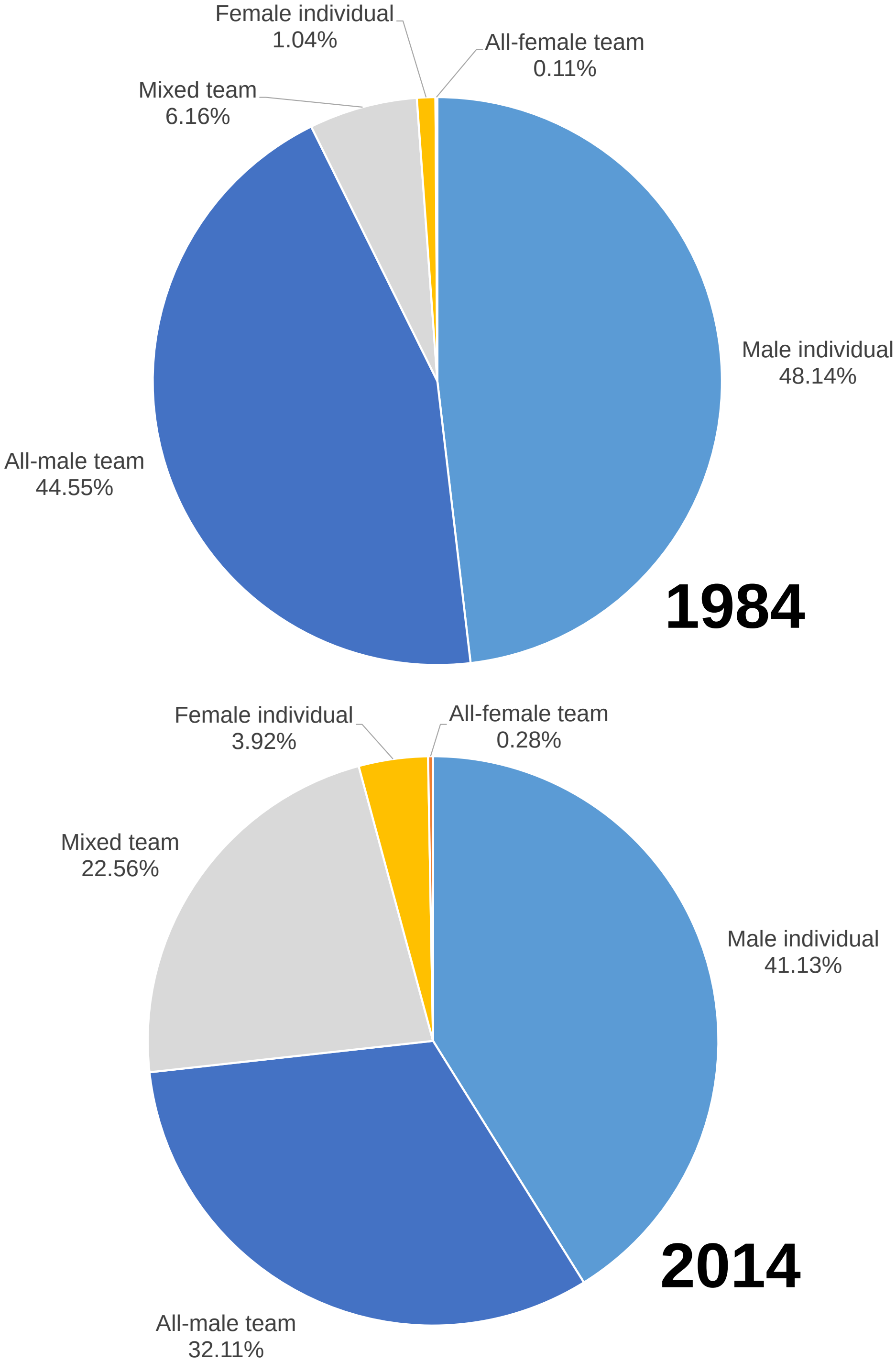



**Figure 7: Comparing the inventor types on patents filed in 1984 and 2014**

Figure 8 shows how PATSTAT patents from the top inventor countries appear once split by the different inventor types. For example, over 21% of patents from French inventors have at least one female inventor (2.0% female individuals, 0.7% all-female teams and 19.0% mixed teams). There is a noticeable difference in the proportion of male individual inventors compared to all-male teams on patents from Korean, Chinese and Taiwanese inventors compared to the other five countries presented, all of which have relatively similar ratios. However it is unclear if this is a real truth or an artefact of the low gender inference rates for Korean, Chinese and Taiwanese inventors as shown previously in Table 6.

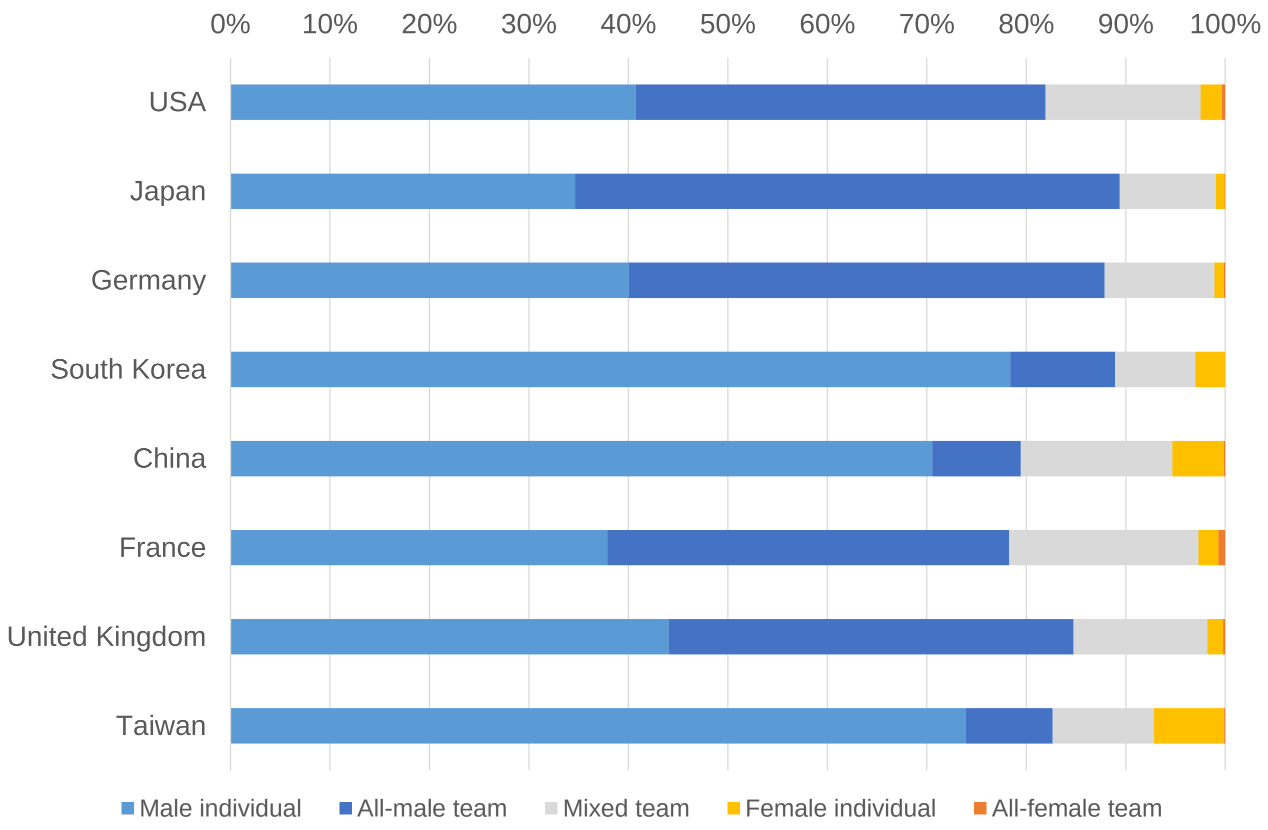



**Figure 8: Patents from top inventor countries split by inventor type**

## 4.5 Technology focus

All patents are classified according to the International Patent Classification (IPC)[42]. Patents are classified based upon the technical features of the invention and therefore provide an insight into the area of technology for which protection is sought. Table 7 shows the inventor gender ratio for the top 20 IPC subclasses with the most inventors in PATSTAT. The average across all IPC subclasses in PATSTAT is 8.3% and Table 7 shows that eight of the top 20 IPC subclasses with the most inventors have an above-average proportion of female inventors and five of the top 10 have over 12% female inventors.

**Table 7: Inventor gender ratio for top 20 IPC subclasses**

| IPC subclass | | Male inventors | Female inventors |
|---|---|---|---|
| A61K | Medical Or Veterinary Science; Hygiene -> Preparations For Medical, Dental, Or Toilet Purposes | 82.8% | 17.2% |
| C07D | Organic Chemistry -> Heterocyclic Compounds | 87.6% | 12.4% |
| A61P | Medical Or Veterinary Science; Hygiene -> Therapeutic Activity Of Chemical Compounds Or Medicinal Preparations | 83.5% | 16.5% |
| H01L | Basic Electric Elements -> Semiconductor Devices; Electric Solid State Devices Not Otherwise Provided For | 95.4% | 4.6% |
| C07C | Organic Chemistry -> Acyclic Or Carbocyclic Compounds | 90.8% | 9.2% |
| G06F | Computing; Calculating; Counting -> Electric Digital Data Processing | 93.8% | 6.2% |
| H04N | Electric Communication Technique -> Pictorial Communication, *e.g.* Television | 95.8% | 4.2% |
| G01N | Measuring; Testing -> Investigating Or Analysing Materials By Determining Their Chemical Or Physical Properties | 87.9% | 12.1% |
| C12N | Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation Or Genetic Engineering -> Micro-Organisms Or Enzymes; Compositions Thereof; Propagating, Preserving, Or Maintaining Micro-Organisms; Mutation Or Genetic Engineering; Culture Media | 79.0% | 21.0% |
| H04L | Electric Communication Technique -> Transmission Of Digital Information, *e.g.* Telegraphic Communication | 94.5% | 5.5% |
| C07K | Organic Chemistry -> Peptides | 80.4% | 19.6% |
| C08L | Organic Macromolecular Compounds; Their Preparation Or Chemical Working-Up; Compositions Based Thereon -> Compositions Of Macromolecular Compounds | 92.7% | 7.3% |
| G11B | Information Storage -> Information Storage Based On Relative Movement Between Record Carrier And Transducer | 96.9% | 3.1% |
| B01J | Physical Or Chemical Processes Or Apparatus In General -> Chemical Or Physical Processes, *e.g.* Catalysis, Colloid Chemistry; Their Relevant Apparatus | 91.8% | 8.2% |
| C08F | Organic Macromolecular Compounds; Their Preparation Or Chemical Working-Up; Compositions Based Thereon -> Macromolecular Compounds Obtained By Reactions Only Involving Carbon-To-Carbon Unsaturated Bonds | 91.9% | 8.1% |
| A61B | Medical Or Veterinary Science; Hygiene -> Diagnosis; Surgery; Identification | 93.0% | 7.0% |
| H01M | Basic Electric Elements -> Processes Or Means, *e.g.* Batteries, For The Direct Conversion Of Chemical Energy Into Electrical Energy | 94.1% | 5.9% |
| B01D | Physical Or Chemical Processes Or Apparatus In General -> Separation | 94.3% | 5.7% |
| B29C | Working Of Plastics; Working Of Substances In A Plastic State In General -> Shaping Or Joining Of Plastics; Shaping Of Substances In A Plastic State, In General; After-Treatment Of The Shaped Products, *e.g.* Repairing | 96.6% | 3.4% |
| A01N | Agriculture; Forestry; Animal Husbandry; Hunting; Trapping; Fishing -> Preservation Of Bodies Of Humans Or Animals Or Plants Or Parts Thereof; Biocides, *e.g.* As Disinfectants, As Pesticides Or As Herbicides; Pest Repellents Or Attractants; Plant Growth Regulators | 88.3% | 11.7% |

[42] https://www.gov.uk/government/publications/patent-classification/patent-classification

Table 8 shows the top 10 and bottom 10 IPC subclasses[43] for female inventors in PATSTAT. There are clear, and somewhat gender stereotypical, differences between the top and bottom IPC subclasses.

### Table 8: Female inventors by IPC subclass

| | IPC subclass | | Male inventors | Female inventors |
|---|---|---|---|---|
| Top 10 | A41C | Wearing Apparel -> Corsets; Brassieres | 62.8% | 37.2% |
| | A61Q | Medical Or Veterinary Science; Hygiene -> Use Of Cosmetics Or Similar Toilet Preparations | 76.6% | 23.4% |
| | A41B | Wearing Apparel -> Shirts; Underwear; Baby Linen; Handkerchiefs | 77.2% | 22.8% |
| | C12N | Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation Or Genetic Engineering -> Micro-Organisms Or Enzymes; Compositions Thereof; Propagating, Preserving, Or Maintaining Micro-Organisms; Mutation Or Genetic Engineering; Culture Media | 79.0% | 21.0% |
| | A47D | Furniture; Domestic Articles Or Appliances; Coffee Mills; Spice Mills; Suction Cleaners In General -> Furniture Specially Adapted For Children | 79.6% | 20.4% |
| | C12Q | Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation Or Genetic Engineering -> Measuring Or Testing Processes Involving Enzymes Or Micro-Organisms; Compositions Or Test Papers Therefor; Processes Of Preparing Such Compositions; Condition-Responsive Control In Microbiological Or Enzymological Processes | 79.6% | 20.4% |
| | A21D | Baking; Equipment For Making Or Processing Doughs; Doughs For Baking -> Treatment, E.G. Preservation, Of Flour Or Dough For Baking, *e.g.* By Addition Of Materials; Baking; Bakery Products; Preservation Thereof | 79.8% | 20.2% |
| | A41G | Wearing Apparel -> Artificial Flowers; Wigs; Masks; Feathers | 80.0% | 20.0% |
| | A01H | Agriculture; Forestry; Animal Husbandry; Hunting; Trapping; Fishing -> New Plants Or Processes For Obtaining Them; Plant Reproduction By Tissue Culture Techniques | 80.1% | 19.9% |
| | C07K | Organic Chemistry -> Peptides | 80.4% | 19.6% |
| Bottom 10 | F02B | Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants -> Internal-Combustion Piston Engines; Combustion Engines In General | 98.1% | 1.9% |
| | D03C | Weaving -> Shedding Mechanisms; Pattern Cards Or Chains; Punching Of Cards; Designing Patterns | 98.1% | 1.9% |
| | F02N | Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants -> Starting Of Combustion Engines; Starting Aids For Such Engines, Not Otherwise Provided For | 98.2% | 1.8% |
| | B61B | Railways -> Railway Systems; Equipment Therefor Not Otherwise Provided For | 98.2% | 1.8% |
| | F02D | Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants -> Controlling Combustion Engines | 98.2% | 1.8% |
| | F01L | Machines Or Engines In General; Engine Plants In General; Steam Engines -> Cyclically Operating Valves For Machines Or Engines | 98.3% | 1.7% |
| | G21B | Nuclear Physics; Nuclear Engineering -> Fusion Reactors | 98.3% | 1.7% |
| | B60T | Vehicles In General -> Vehicle Brake Control Systems Or Parts Thereof; Brake Control Systems Or Parts Thereof, In General; Arrangement Of Braking Elements On Vehicles In General; Portable Devices For Preventing Unwanted Movement Of Vehicles; Vehicle Modifications To Facilitate Cooling Of Brakes | 98.4% | 1.6% |
| | G10B | Musical Instruments; Acoustics -> Organs; Harmoniums | 98.6% | 1.4% |
| | F02P | Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants -> Ignition, Other Than Compression Ignition, For Internal-Combustion Engines; Testing Of Ignition Timing In Compression-Ignition Engines | 98.6% | 1.4% |

Table 9 shows a subset of the data presented in Table 8 and highlights the top 10 and bottom 10 IPC subclasses for female inventors in PATSTAT over the last 15 years. It is

[43] With over 600 IPC subclasses, analysis was limited to subclasses comprising over 1000 inventors

clear from Figure 2 that the proportion of female inventors has increased rapidly over the last 40 years and, since the gender demographic within industry in the 21st Century is very different to what it was for most of the 20th Century, it could be considered to be misrepresentative and potentially misleading to read too much into the overall data presented in Table 8, notably the number of traditional associations about female inventors. The data in Table 9 however appears to suggest that this is not the case; although the proportion of female inventors is slightly higher (but not unexpected) in the top 10 IPC subclasses, the clear, and somewhat gender stereotypical, differences between the top and bottom IPC subclasses are very similar between data limited to the last 15 years and data from all patents in PATSTAT dating back to 1782.

**Table 9: Female inventors by IPC subclass, 2000-2015**

| IPC subclass | | | Male inventors | Female inventors |
|---|---|---|---|---|
| **Top 10** | A41C | Wearing Apparel -> Corsets; Brassieres | 58.4% | 41.6% |
| | A61Q | Medical Or Veterinary Science; Hygiene -> Use Of Cosmetics Or Similar Toilet Preparations | 72.6% | 27.4% |
| | A41B | Wearing Apparel -> Shirts; Underwear; Baby Linen; Handkerchiefs | 73.0% | 27.0% |
| | A41F | Wearing Apparel -> Garment Fastenings; Suspenders | 75.3% | 24.7% |
| | B68B | Saddlery; Upholstery -> Harness; Devices Used In Connection Therewith; Whips Or The Like | 75.4% | 24.6% |
| | A41G | Wearing Apparel -> Artificial Flowers; Wigs; Masks; Feathers | 76.2% | 23.8% |
| | B68C | Saddlery; Upholstery -> Saddles; Stirrups | 76.4% | 23.6% |
| | A21D | Baking; Equipment For Making Or Processing Doughs; Doughs For Baking -> Treatment, *e.g.* Preservation, Of Flour Or Dough For Baking, *e.g.* By Addition Of Materials; Baking; Bakery Products; Preservation Thereof | 76.7% | 23.3% |
| | C07K | Organic Chemistry -> Peptides | 76.9% | 23.1% |
| | C12N | Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation Or Genetic Engineering -> Micro-Organisms Or Enzymes; Compositions Thereof; Propagating, Preserving, Or Maintaining Micro-Organisms; Mutation Or Genetic Engineering; Culture Media | 77.2% | 22.8% |
| **Bottom 10** | B60W | Vehicles In General -> Conjoint Control Of Vehicle Sub-Units Of Different Type Or Different Function; Control Systems Specially Adapted For Hybrid Vehicles; Road Vehicle Drive Control Systems For Purposes Not Related To The Control Of A Particular Sub-Unit | 97.6% | 2.4% |
| | F01C | Machines Or Engines In General; Engine Plants In General; Steam Engines -> Rotary-Piston Or Oscillating-Piston Machines Or Engines | 97.8% | 2.2% |
| | F02D | Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants -> Controlling Combustion Engines | 97.9% | 2.1% |
| | B60T | Vehicles In General -> Vehicle Brake Control Systems Or Parts Thereof; Brake Control Systems Or Parts Thereof, In General; Arrangement Of Braking Elements On Vehicles In General; Portable Devices For Preventing Unwanted Movement Of Vehicles; Vehicle Modifications To Facilitate Cooling Of Brakes | 97.9% | 2.1% |
| | D03C | Weaving -> Shedding Mechanisms; Pattern Cards Or Chains; Punching Of Cards; Designing Patterns | 97.9% | 2.1% |
| | F01L | Machines Or Engines In General; Engine Plants In General; Steam Engines -> Cyclically Operating Valves For Machines Or Engines | 97.9% | 2.1% |
| | F02N | Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants -> Starting Of Combustion Engines; Starting Aids For Such Engines, Not Otherwise Provided For | 98.0% | 2.0% |
| | B25D | Hand Tools; Portable Power-Driven Tools; Handles For Hand Implements; Workshop Equipment; Manipulators -> Percussive Tools | 98.2% | 1.8% |
| | G10B | Musical Instruments; Acoustics -> Organs; Harmoniums | 98.2% | 1.8% |
| | F02P | Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants -> Ignition, Other Than Compression Ignition, For Internal-Combustion Engines; Testing Of Ignition Timing In Compression-Ignition Engines | 98.5% | 1.5% |

The radar map in Figure 9 shows the thematic specialisation of male and female inventors according to the eight IPC sections. There is a high specialisation of female inventors in sections A (Human Necessities, 36%) and C (Chemistry; Metallurgy, 34%). This contrasts with male inventors who are more evenly spread across sections A, B, C, G and H, which range from 15% in section B (Performing Operations; Transporting) to 24% in section C (Chemistry; Metallurgy).

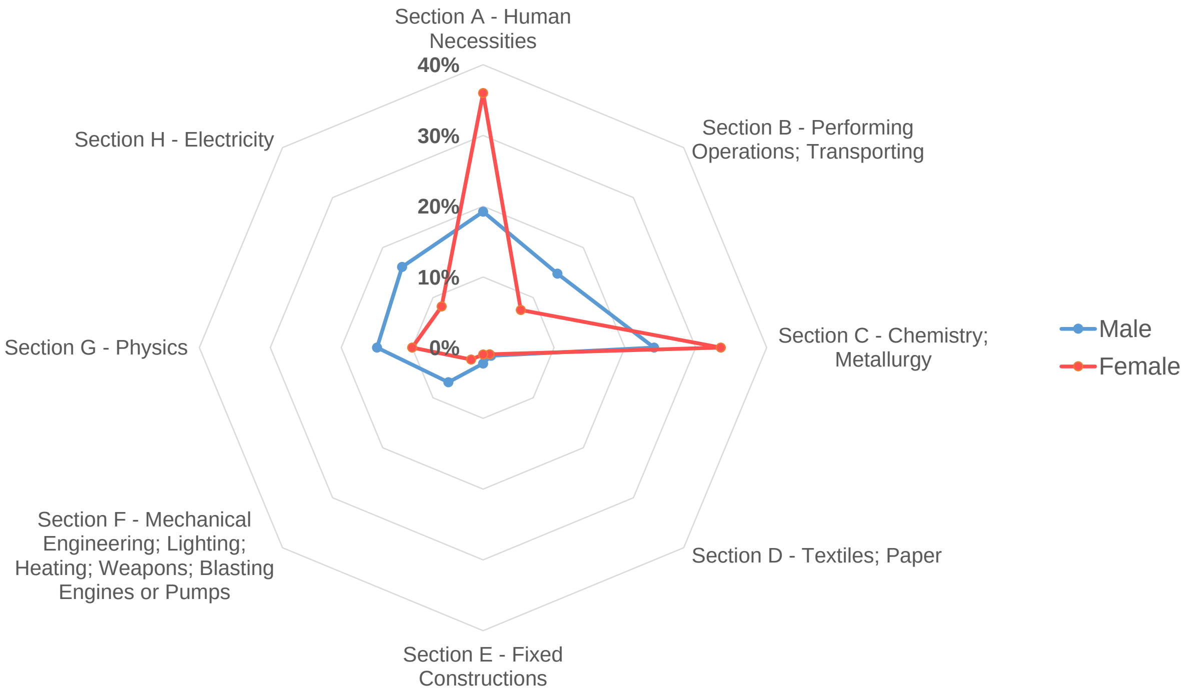



**Figure 9: Thematic specialisation of male and female inventors by IPC section**

The granularity of IPC subclasses shown in Table 7 to Table 9 means that outliers may be exaggerated, especially as each IPC subclass is not equal in size. Previous work by the World Intellectual Property Organisation (WIPO) has produced a technology concordance table[44] that links the IPC symbols with five technology sectors sub-divided into 35 fields of technology. This provides a good level of granularity for this study and Table 10 shows the inventor gender ratio by WIPO technology field. For female inventors, there is a clear divide between the chemistry technology fields and the mechanical and electrical engineering fields with six of the 11 chemistry fields falling in the top seven fields for female inventors.

[44] http://www.wipo.int/ipstats/en/statistics/technology_concordance.html

**Table 10: Female inventors by WIPO technology concordance fields**

| WIPO technology field | Male inventors | Female inventors |
|---|---|---|
| Chemistry: Biotechnology | 80.2% | 19.8% |
| Instruments: Analysis of biological materials | 81.8% | 18.2% |
| Chemistry: Pharmaceuticals | 83.6% | 16.4% |
| Chemistry: Food chemistry | 83.7% | 16.3% |
| Chemistry: Organic fine chemistry | 87.1% | 12.9% |
| Chemistry: Basic materials chemistry | 89.7% | 10.3% |
| Chemistry: Micro-structural and Nanotechnology | 90.2% | 9.8% |
| Instruments: Medical technology | 90.9% | 9.1% |
| Electrical engineering: IT methods for management | 91.6% | 8.4% |
| Other fields: Other consumer goods | 91.7% | 8.3% |
| Chemistry: Macromolecular chemistry, polymers | 92.1% | 7.9% |
| Chemistry: Chemical engineering | 93.4% | 6.6% |
| Other fields: Furniture, games | 93.5% | 6.5% |
| Chemistry: Materials, metallurgy | 93.9% | 6.1% |
| Electrical engineering: Digital communication | 94.0% | 6.0% |
| Chemistry: Surface technology, coating | 94.1% | 5.9% |
| Mechanical engineering: Textile and paper machines | 94.1% | 5.9% |
| Chemistry: Environmental technology | 94.2% | 5.8% |
| Electrical engineering: Computer technology | 94.2% | 5.8% |
| Instruments: Measurement | 95.1% | 4.9% |
| Mechanical engineering: Other special machines | 95.2% | 4.8% |
| Instruments: Control | 95.3% | 4.7% |
| Electrical engineering: Semiconductors | 95.4% | 4.6% |
| Instruments: Optics | 95.5% | 4.5% |
| Electrical engineering: Telecommunications | 95.6% | 4.4% |
| Electrical engineering: Electrical machinery, apparatus, | 95.8% | 4.2% |
| Mechanical engineering: Handling | 95.8% | 4.2% |
| Other fields: Civil engineering | 96.2% | 3.8% |
| Electrical engineering: Audio-visual technology | 96.3% | 3.7% |
| Mechanical engineering: Thermal processes and apparatus | 96.5% | 3.5% |
| Mechanical engineering: Machine tools | 96.7% | 3.3% |
| Mechanical engineering: Transport | 97.2% | 2.8% |
| Electrical engineering: Basic communication processes | 97.2% | 2.8% |
| Mechanical engineering: Mechanical elements | 97.3% | 2.7% |
| Mechanical engineering: Engines, pumps, turbines | 97.5% | 2.5% |

As with the IPC analysis (Table 8 and Table 9), if you take a data subset from PATSTAT for patents in the last 15 years (2000-2015), the inventor gender ratio for the top six and bottom eight fields of technology using the WIPO technology concordance table remain the same, albeit with slightly higher proportion of female inventors compared to Table 10 (*i.e.* 22.3% for biotechnology and 12.3% for basic materials chemistry; down to 5% for civil engineering and 3.2% for engines, pumps and turbines).

Figure 10 shows the inventor types across PATSTAT patents once linked to the five WIPO technology sectors. Although 68% of chemistry-related patents come from all-male inventors (23% male individuals and 45% all-male teams), female inventors have the biggest share in this sector with the remaining 32% of chemistry-related patents having at least one female inventor (1.7% female individuals, 0.6% all-female teams and 29.5% mixed teams).

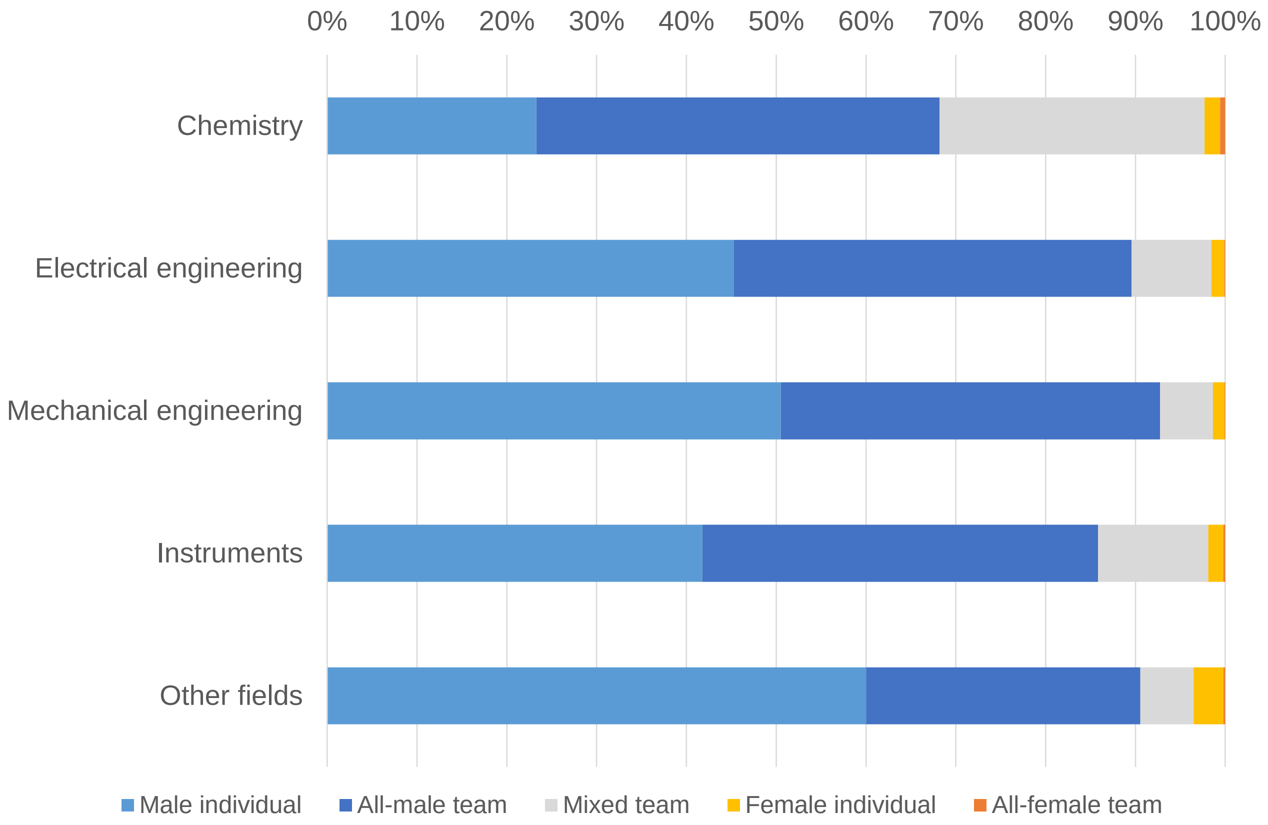



**Figure 10: PATSTAT patents linked to WIPO technology concordance sectors, split by inventor type**

When the data presented in Figure 10 is sub-divided into the 35 WIPO technology fields (as listed in Table 10), certain fields show even higher proportions of female inventorship; for example, 49% of biotechnology-related patents have at least one female inventor (1.9% female individuals, 0.9% all-female teams and 46.3% mixed teams) and 45% of pharmaceutical-related patents have at least one female inventor (2.2% female individuals, 0.7% all-female teams and 41.9% mixed teams). At the other end of the scale, only 5.4% of engine-, pump- and turbine-related patents have at least one female inventor (0.8% female individuals, 0.02% all-female teams and 4.5% mixed teams)[45].

[45] All data can be downloaded from https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship

# 5 Discussion

The results of this study looking at disaggregating the gender of inventors on all published patents worldwide are not dissimilar from the findings of the earlier preliminary study that was limited to GB patent applications. This is not that surprising because there is plenty of anecdotal evidence across the world that levels of female patenting are low. Only 7% of all inventors worldwide are female and, although this is slowly increasing, in recent years it still remains relatively low at around 10%. There are some differences between inventors from different countries and across different technical disciplines but there are many other social and economic factors that raw quantitative patent analysis does not take into account. The complete findings and outputs of this study have been published[46] so that it may inspire future research projects to combine both quantitative and qualitative data to provide a richer analysis of female inventorship across the world.

Restrained with time and resources, this study is not without its limitations and caveats. As with the previous research in this field, the accuracy of gender-based patent analysis is only as good as the quality of name-gender assignments in the underlying master dataset that the patent data is matched to. This study used multiple data sources for this, which were further enhanced using manual cleaning and country-specific dictionaries. These data sources[47] are inherently biased towards 'western' (*e.g.* non-Asian) first names and so a lot of the manual cleaning undertaken in this study was performed to try to improve the gender inference rates for Asian inventors. Table 6 shows the gender inference rates for the top inventor countries and it is clear that specific knowledge of the 'baby naming' system for Korea, China and Taiwan could significantly improve the match rates for these countries, as well as other inventor countries with poor match rates. Similarly, country-specific data dictionaries have only been implemented where known (*e.g.* for names such as Jean, Andrea, Nicola etc) and could also be improved with local knowledge. Further improvements could also be made by studying surname data since in certain countries (*e.g.* Russia) it is possible to infer gender from surnames.

As mentioned above, this study would benefit from further work to improve the match rates for Asian inventors. There are certain limitations in the inventor name data contained within PATSTAT that significantly impact how much these match rates could actually be improved. Patent data contained in PATSTAT is provided to the EPO by each national or regional patent office and this affects the cleanliness and quality of this data, in particular regarding machine-transliterations of non-Latin characters. For example, non-Latin names such as Chinese names may not be accurately transliterated when a Chinese inventor files for a patent in Australia and, once mistransliterated, it is highly unlikely that there will be a successful match with the name-gender dataset. One possible option to improve gender inference rates could be to use PATSTAT patent family data to see if cleaner inventor name data exists for equivalent patent applications in another jurisdiction.

[46] https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship

[47] See section 3.2

It was also not within the scope of this research to investigate gender migration and further match rate improvements could be obtained by considering names that have shifted gender over time. For example, in the master name-gender assignment dataset used in this study the name *Meredith* is considered to be 94% female and, since this research required set a 95% certainty threshold, any inventors with the name *Meredith* would not be matched. According to *nameberry.com*[48], *Meredith* was 100% male in 1883, and 100% female in 2012 with the gender line crossed permanently in 1921. In this case it appears that the name *Meredith* has migrated gender over time. Further research could theoretically enable inventors with the name *Meredith* to be assigned a gender with a 95% certainty depending on the era in which the patent was filed[49].

The analysis contained in this report provides a brief overview of the type of patent analysis that can be undertaken using the UK IPO inferred inventor gender table. It is believed however that the real value in this add-on table for PATSTAT can only be realised when specific questions are asked and the full breadth of PATSTAT is queried alongside inferred inventor gender data. Any further analysis should bear in mind the nuances of the patent system and patent data, and some of the common pitfalls of analysing patent data, as explained in *The Patent Guide*[40]. For example, this report does not feature any analysis of first named inventors (primary inventors) which is used in some of the analysis undertaken in the previous research discussed in section 2; although there is a legal requirement for all inventors to be listed on a patent application, there is no legal requirement for inventors to be listed in any specific order[50] so no meaningful patent analysis can be undertaken by studying primary inventors.

Additionally, this study has not performed any analysis focussing on the corresponding applicant/assignee type (*e.g.* corporate, academia, government) associated with each inventor on each patent application. Although possible, care should be taken when undertaking this type of analysis because it is only realistically possible when there is a single named applicant. If there are multiple inventors and multiple applicants (*i.e.* collaborating co-applicants) it will not be possible to assign which inventor is employed by which applicant[51]. Anecdotal evidence suggests higher levels of female inventorship within academic than within industry and, although beyond the scope of this study, further work could be undertaken to draw these additional links using the inventor gender inference table.

Whilst absolute patent counts do not give a direct measure of innovation, they are well known as a measurable ‘output’ of STEM industries. This study has shown it is now possible to use worldwide patent data as a good source of evidence to inform the wider gender debate. Further studies could focus on linking this data to other data sources, such as the well-recognised and measurable ‘inputs’ of STEM industries to form a bigger picture, to provide a sound basis for evidence-based policy within government and industry.

---

48 *http://nameberry.com/blog/unisex-baby-names-names-that-morphed-from-blue-to-pink*

49 Although large error margins would be required since inventor age data is not available

50 The same applies to patent applicants/assignees

51 Over 12.5% of PATSTAT patents with applicant information have co-applicants, *i.e.* more than one named applicant

# 6 Conclusions

This research aimed to infer the gender of inventors on all patents worldwide. The UK IPO is not aware of the existence of any similar previous work on this scale and the European Patent Office (EPO) Worldwide Patent Statistics database, PATSTAT, has been used to include all published patents worldwide dating back to the $18^{th}$ Century. An inventor gender inference add-on table for PATSTAT has been produced where the gender of over 70% of all PATSTAT inventor identifiers have now been inferred with a high degree of confidence.

The inventor gender inference table can be linked with other PATSTAT tables and it is now possible to perform worldwide gender-based patent analysis for the first time. Alongside this report, the inventor gender inference table has also been published in the hope that it will inspire other academics and patent analysts to undertake future research projects to provide deeper insights into female inventorship across the world.

Subsequent statistical analysis about the patenting activity of female inventors worldwide provided some quantitative data to back up the anecdotal evidence within the IP industry about the gender gap. For several decades within the $20^{th}$ Century women represented less than 2% of inventors on published patent applications worldwide but this has steadily risen to over 10% in recent years. Although absolute numbers remain relatively low, the last 20 years has seen a 100% increase in the proportion of female inventors. The analysis shows that the gender gap in worldwide patenting is substantial but it is decreasing.

There are some noticeable differences in the inventor gender ratio worldwide. Notwithstanding the numerous socio-economic factors involved, the patent data reveals a stark contrast in the proportion of female inventors, with French (11.7%) and Russian (15.7%) female inventors a long way ahead of Japanese (3.7%), Korean (4.4%) and German (5.5%) female inventors. British (7.3%) and American (8.7%) female inventors are relatively close to the worldwide average of 7.2%.

The overall proportion of patents involving a female inventor (either working alone or as part of a team) has increased by more than 500% between 1975 and 2015 and in recent years over 25% of all patents have at least one named female inventor. The number of all-female teams however remains very low with only 0.3% of patents coming from all-female teams over the last 10 years. Although historical analysis reveals ever-increasing levels of female patenting, the growth rate is slow and the absolute numbers are still low. The world of patenting remains male-dominated and even in 2014 there is a clear gender disparity with 73% of all worldwide patent applications coming from all-male inventors, rising to almost 96% when mixed teams are considered (*i.e.* 96% of all patent applications worldwide in 2014 had at least one male inventor).

The demographic profile of inventors is also shown to vary substantially by technology. Analysing the granular classification codes applied to each patent application revealed a number of traditional associations with the highest proportions of female inventors listed on patents relating to brassieres, clothing and cosmetics, and the lowest proportions on patents relating to combustion engines and vehicles. A higher level analysis using WIPO's technology concordance table smoothed out the niche technical fields with lower absolute level of patenting. It revealed a more accurate reflection of the STEM industries in which the most women are employed, with several chemistry areas including biotechnology, food chemistry and pharmaceuticals having the highest proportion of female inventors.

These conclusions show a pattern that is consistent with previous research but for the first time it is possible to analyse gender profiles in patenting worldwide without the data coverage limitations of the previous studies. Further work is required to continue to improve data coverage and gender inference rates, notably on inventor names transliterated from non-Latin alphabets or via cross-validation using patent family information. Nonetheless, this study is considered to be an important step towards a better understanding of gender differences in patenting and the worldwide inventor gender inference table is believed to be a new analysis tool to provide a sound basis for future evidence-based policy within government and industry.

Concept House
Cardiff Road
Newport
NP10 8QQ

**Tel**: 0300 300 2000
**Fax**: 01633 817 777
**Email**: information@ipo.gov.uk
**Web**: www.gov.uk/ipo

**Facebook**: TheIPO.UK
**Twitter**: @The_IPO
**YouTube**: ipogovuk

For copies in alternative formats please contact our Information Centre.

**When you no longer need this booklet, please recycle it.**



DPS-004404