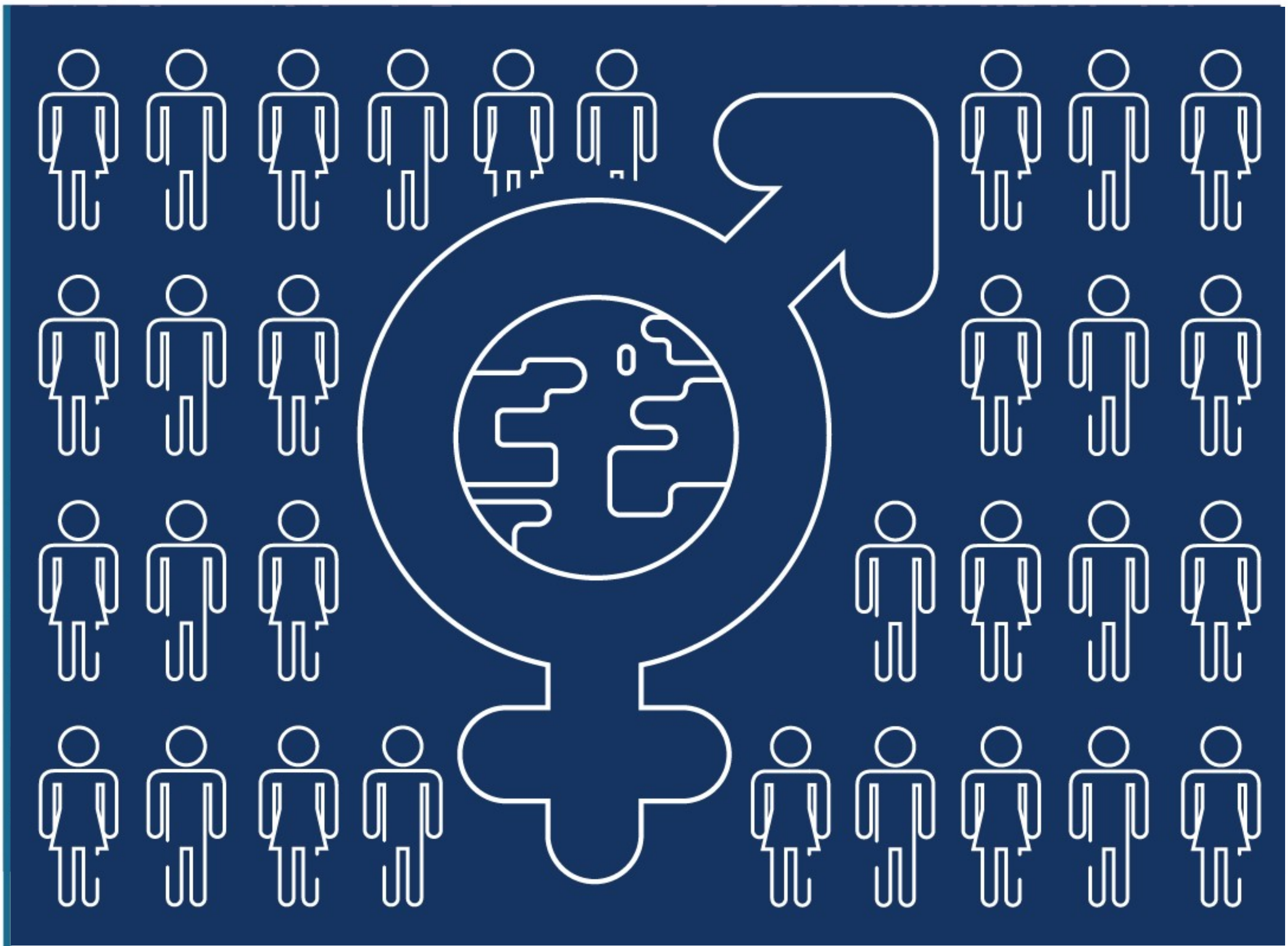
Intellectual Property Office

Gender profiles in worldwide patenting

An analysis of female inventorship (2019 edition)



IPO | Making life better through IP

Intellectual Property Office is an operating name of the Patent Office

2019/09

ISBN 978-1-910790-75-5
Gender profiles in worldwide patenting: an analysis of female inventorship (2019 edition)

This report was prepared by the Economics, Research and Evidence team at the Intellectual Property Office, September 2019.

Any enquiries regarding this publication should be sent to:

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

Tel: 0300 300 2000 Fax: 01633 817 777

e-mail: statistics@ipo.gov.uk

This publication is available from our website at https://www.gov.uk/ipo

Contents

Executive Summary

The representation of women within STEM fields (science, technology, engineering and mathematics) is of great importance to governments and policymakers. Outputs of work undertaken within STEM include inventions for which protection is sought through patent applications. Consequently, a recurring question that has been asked of the IPO of the UK over the past few years relates to patent statistics about female inventors. However, the gender of inventors is generally not disclosed during the patent application process. This report is an experimental study into the representation of women in worldwide patent inventorship, which is based on inferences made based on bibliographic data that includes the forenames of inventors.

An increasing proportion of patent inventors worldwide are female. Between 1998 and 2017, the proportion of female inventors worldwide has almost doubled from 6.8% to 12.7%. Further, the proportion of patent applications that name a female amongst their inventors rose from 12% to 21% over the same period, and the proportion of applications with at least as many female inventors as males rose from 3% to 8%.

The extent to which women participate in patent inventorship has been analysed. The majority of patents with a female inventor comprise one woman, either working as a lone inventor or as part of a male-dominated team. However, the proportion of patent applications which have several female inventors has more than doubled, from 2% in 1998 to 4.5% in 2017.

There are different trends in female inventorship depending on the technology area. The technology areas that enjoy the greatest levels of female participation are Biotechnology, Pharmaceuticals and Organic Chemistry. Women have contributed to the inventions of approximately half of the patent applications made in these fields between 1998-2017.

This study has provided a dataset linking inventors to their inferred genders, and so may represent a good source of evidence to inform the wider gender debate. Further studies could focus on linking this dataset 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.

Introduction

Background

In the UK, the Government Office for Science[1], supported by organisations such as Innovate UK[2], the UKRI[3], the Royal Academy of Engineering[4], the IET[5] and campaigns such as WISE[6] 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[7] highlight the impact of their work within the education sector with the number of females attaining STEM higher level qualifications increasing. The same surveys also highlight the gender demographic transition to the workplace with women making up only 23% of the STEM workforce in 2017 and women accounting for only 11% 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. Much less 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 IPO of the UK over the past few years 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. An analysis of inventor information would therefore help to understand the representation of males and females within the patent system. To place this work in context, Figure 1 shows the progression of females and males through higher education and then on to careers as researchers and inventors. Progressively lower proportions of females reach each step; this is often referred to as the ‘leaky pipeline’. The solid lines in Figure 1 are taken from a summary of a recent report prepared by UNESCO[8], and tracks the proportion of females in higher education and undertaking research. The proportion of females that go on to apply for patents as inventors is a natural next step in this progression. The dotted line summarises the findings of the present study and indicates that the proportion of inventors that are female is lower still than the proportion of females undertaking research. Throughout this study, various cross-sections of patent data are analysed to assist in understanding the underlying causes of female under-representation in patent inventorship.

---

[1] GO-Science (Government Office for Science)https://www.gov.uk/government/organisations/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] UK Research and Innovation- http://www.ukri.org

[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] For example, WISE: UK Statistics (https://www.wisecampaign.org.uk/statistics/) and IET, Women in STEM: Statistics and facts (https://communities.theiet.org/files/7976#.VbTQ7fkbJ_8)

[8] UNESCO Science Report: towards 2030, published in 2015. Infographic available from https://en.unesco.org/sites/default/files/usr2015_kakemonos_gender_en.pdf

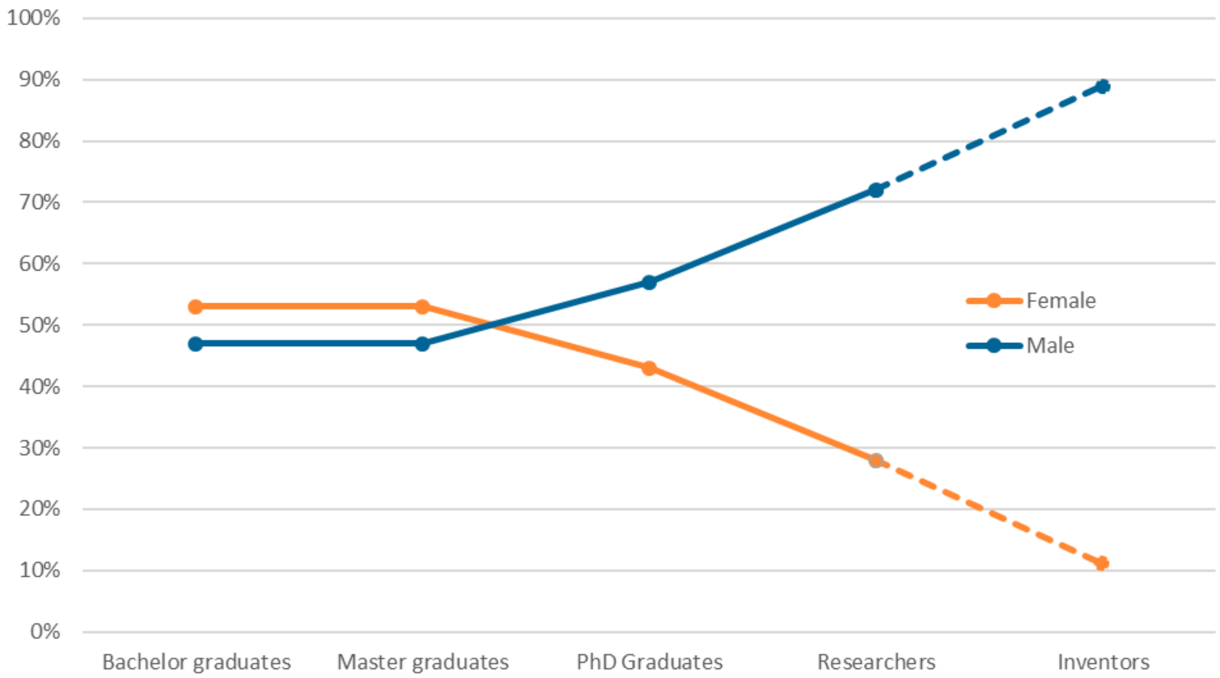


Figure 1: The 'leaky pipeline' - the proportion of females and males reaching each stage of higher education, further research and inventorship

The extent of female participation in inventorship has been analysed only recently, and relies on patent bibliographic data. An early contribution, the contents of which are developed further in this report, was made by the IPO of the UK in 2016[9]. Later that year, the World Intellectual Property Organisation (WIPO) published a working paper[10] in which the gender was inferred for inventors on patent applications made under the Patent Cooperation Treaty (PCT) system. For such applications, information about the nationality and country of residence of inventors is available, and the proportions of female inventors were compared for different nationalities, residence countries, technology types and applicants. Alongside their working paper, WIPO also published a World Gender-Name Dictionary, which was generated through a combination of data sources including native speakers of Chinese, Indian, Japanese and Korean amongst WIPO staff.

More recently, the United States Patents and Trademarks Office published a study[11] of female inventorship associated with patent applications filed with the USPTO. This study analysed trends in female inventorship, comparing inventorship proportions to the proportion of females working in various STEM sectors, analysing differences in female inventorship trends by state, by technology sector, by team size and by applicant.

This report updates the results presented in the earlier report by the IPO of the UK, and extends the analysis to consider broader aspects of female inventorship. This report quantifies how the proportion of female inventorship has increased over the years and how this increase might depend on the subject matter of the patent and the country of residence of the inventor. Patent applications may arise through collaboration between teams of inventors, which may be all-male, all-female, or mixed. The relative proportions of these types of team is also analysed in this report, thus providing insight into the extent to which women contribute towards inventions for which patents are sought.

[9] Gender Profiles in Worldwide Patenting: An analysis of female inventorship, published in 2016. Available at https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship

[10] Martinez et al., "Identifying the gender of PCT inventors". WIPO Economic Research Working Paper No. 33, November 2016. Available at https://www.wipo.int/publications/en/details.jsp?id=4125

[11] USPTO, "Progress and Potential: A profile of women inventors on U.S. patents" Office of the Chief Economist IP Data Highlights, Number 2, February 2019. Available at https://www.uspto.gov/learning-and-resources/ip-policy/economic-research/progress-potential

Methodological features

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 previously been straightforward to provide statistical information about the gender of inventors named on patent applications. Recent name-gender inference work by several academic researchers[12,13] may be adapted to infer information about inventors based on their names as recorded by patent offices. This report relies on bibliographic information available from the Spring 2019 Edition of PATSTAT[14], which collates information provided by many patent offices. The data sources used to infer gender from forenames are inherently biased towards 'western' 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 other inventors, such as those with Asian names. To further improve coverage, an iterative approach (similar to that taken by WIPO in their working paper[10]) was taken towards inferring the gender of forenames. Inventors may have multiple forenames, and so this may be used to infer the gender of forenames that are absent from the initial dataset. This relies on an accurate extraction of forenames from the name field, and then on an accurate attribution of those forenames as either male or female. The methodology used to do this is described in detail in the Technical Annex.

The results presented in this report are experimental and rely solely on information provided by inventors via the patent offices at which they file applications. A key assumption is that the gender of an inventor can reliably be inferred based on their forenames. Although care has been taken to ensure that inventor gender is attributed only where confidence is high, attribution errors may nonetheless be present. In particular, no account is taken of non-binary gender as it is impossible to infer this information based solely on forename data.

We are keen to seek views and would welcome your feedback on the contents of this report. If you have any comments, please email statistics@ipo.gov.uk.

---

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

[13] 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

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

Historical analysis of female inventorship

Worldwide trends from 1915 onwards

Historically, females have been underrepresented amongst patent inventors. Figure 2 plots the proportion of female inventors, where each inventor is counted according to the number of patent applications they were credited with each year. Before 1965, the proportion of female inventors was generally between 2% and 3%. Since then, the proportion of female inventors has risen at an accelerating pace, having risen to 6.8% in 1998, and almost doubling since then to reach 12.7% in 2017.

Figure 3 also covers the time period 1915-2017 but focusses instead on female participation in patent inventorship. The proportion of patent applications with at least one female inventor are shown. Also shown are the number of patent applications for which all inventors are female (i.e. a single female inventor, or an all-female team) and the proportion of patent applications for which at least half of inventors are female (i.e. one or both of a two-inventor team are female, or two or more of a four inventor team[15]. This figure also shows that female inventorship has risen dramatically over the past 20 years, with over one in five applications filed in 2017 having at least one female inventor. Between 1980 and 2000, the proportion of patent applications with a female inventor grew from 8% to 13%. Meanwhile, the proportion of applications with at least as many females as males remained steady at around 3%. This suggests that within that period, the majority of female inventorship was as part of larger teams that were mostly male. This proportion of applications has since risen from 3% to 8% in 2017, which indicates that females have recently come to play a larger role in patent inventorship per application.

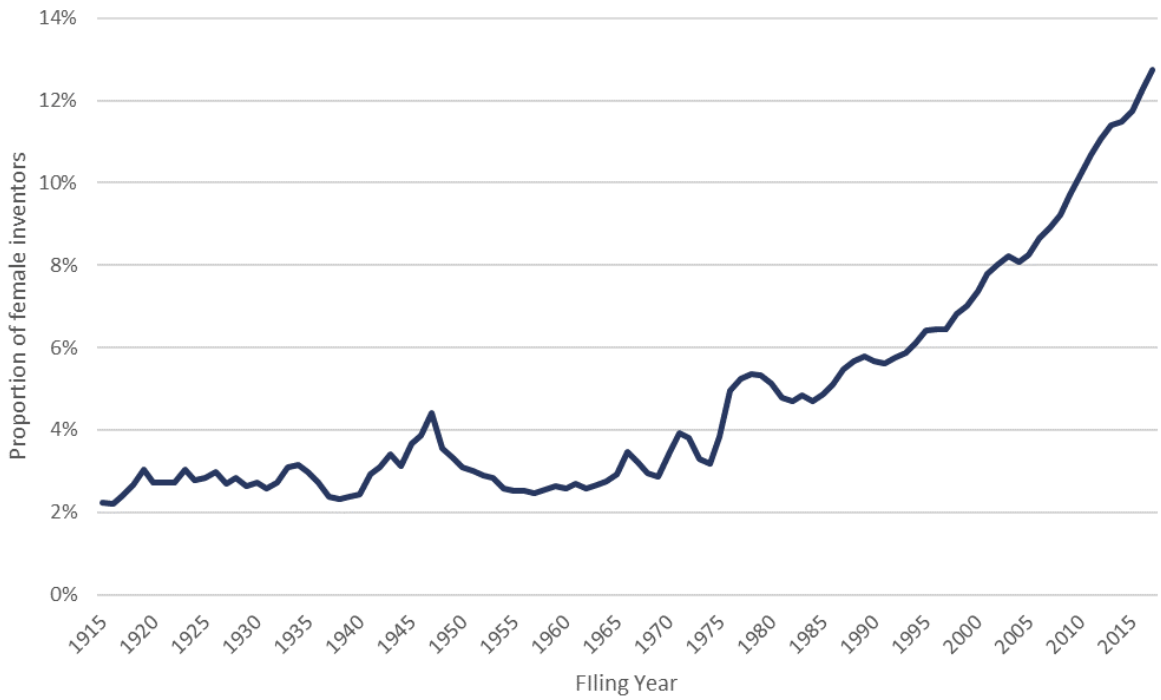


Figure 2: Proportion of female inventors associated with patent applications worldwide, 1915-2017

[15] Proportions of inventors are based on counts of inventors with an inferred gender. For example, a team with two females, two males and one inventor with no inferred gender is considered to comprise 50% (2 out of 4) females

A more detailed analysis of female participation in patent inventorship, either as sole inventors or as members of a team of inventors, is provided later in this report.

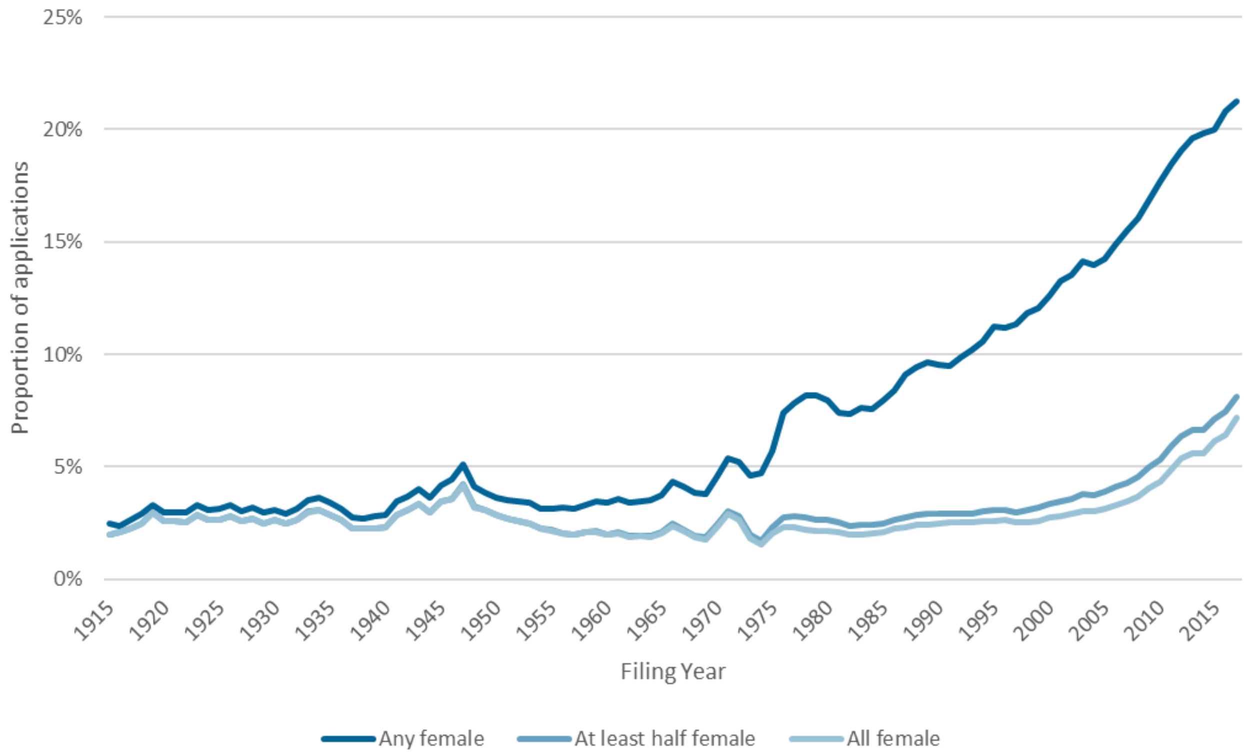


Figure 3: Proportion of worldwide patent applications associated with one or more female inventors, 1915-2017

In light of the sharp rise in female inventorship over the past 20 years, the remainder of this report focusses on the 20-year period 1998-2017. Data for 2018 has not been reported on here. Data for that year will be incomplete as not all applications filed in 2018 will have been published when the Spring 2019 Edition of PATSTAT was released. However, the dataset that accompanies this report[16] includes all available data from the year 2018.

[16] The dataset, and a copy of this report, are available at: https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship-2019-edition

Proportions of female inventorship by country

Figure 4 depicts female inventorship based on self-declared residence information[17]. Of the seven countries with the highest number of resident patent inventors, China shows the largest increase of female inventors over the past 20 years (from 10% in 1998 to 14% in 2017). France has had the largest proportion of female inventors since 1998, and throughout the period 1998-2017 has remained steady at around 16%, which is somewhat higher than the other countries shown. Over the same time period, female inventorship in the UK has risen from 8% to 11%.

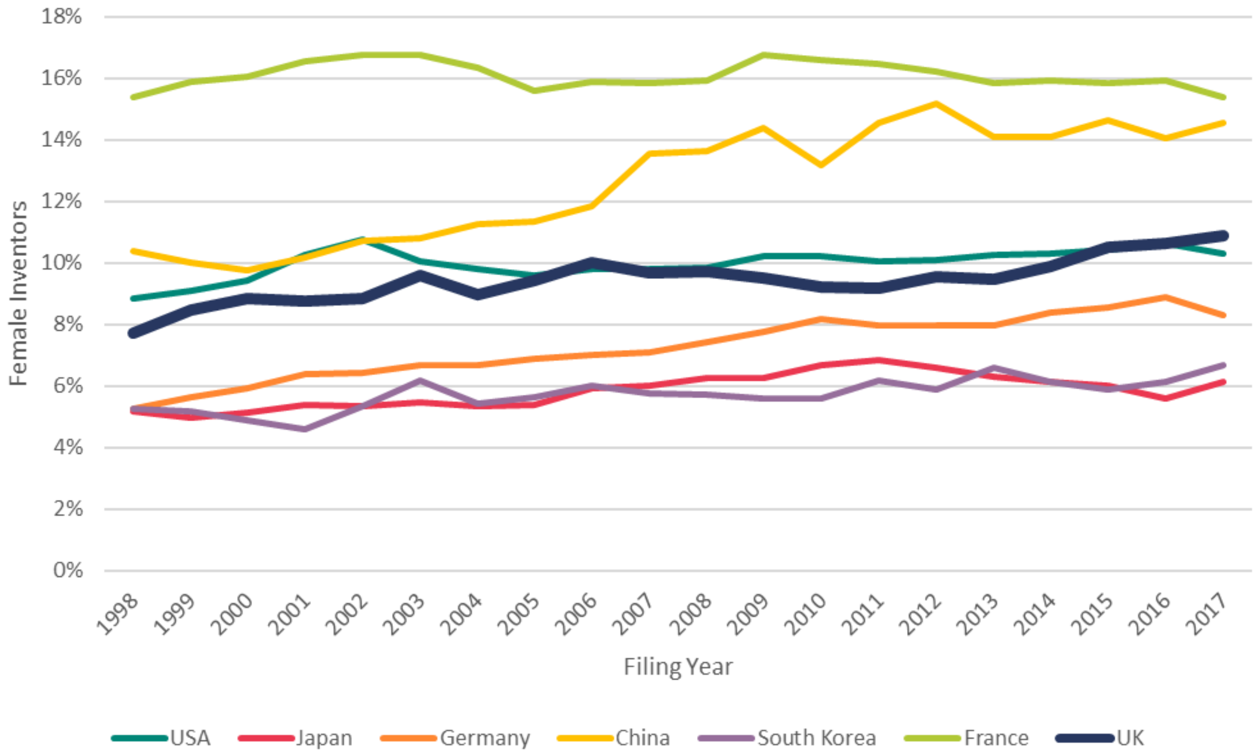


Figure 4: Annual trends in the proportion of female inventors by residence country, 1998-2017

Figure 5 is a choropleth map showing the proportion of female inventors, by declared country of residence, relating to patent applications filed between 1998 and 2017. Results are shown for the 50 countries with the largest number of resident inventors. It should be noted that several countries in South America and Africa, which have low absolute numbers of inventors, have a relatively high percentage of female inventors but are not shown to enable comparisons to be made more easily between the countries shown. Amongst these countries, the largest proportions of female inventors are generally in Asia and Eastern Europe.

Figure 6 represents the same data as Figure 5, but focusses on countries in Europe due to their small geographical size. It shows that a high proportion of inventors from Eastern European countries are female, and it can also be seen that of the Western European countries, a higher proportion of Spanish, French and Italian inventors are female than in the UK, Germany and Scandinavia.

[17] Because declaration of residency is not required by all patent offices, PATSTAT has incomplete coverage, with 19.7 million inventors out of 49.6 million (40%) having no such information. Figure 4 has therefore been generated on the assumption that inventor records with residency information form a representative sample of all inventors

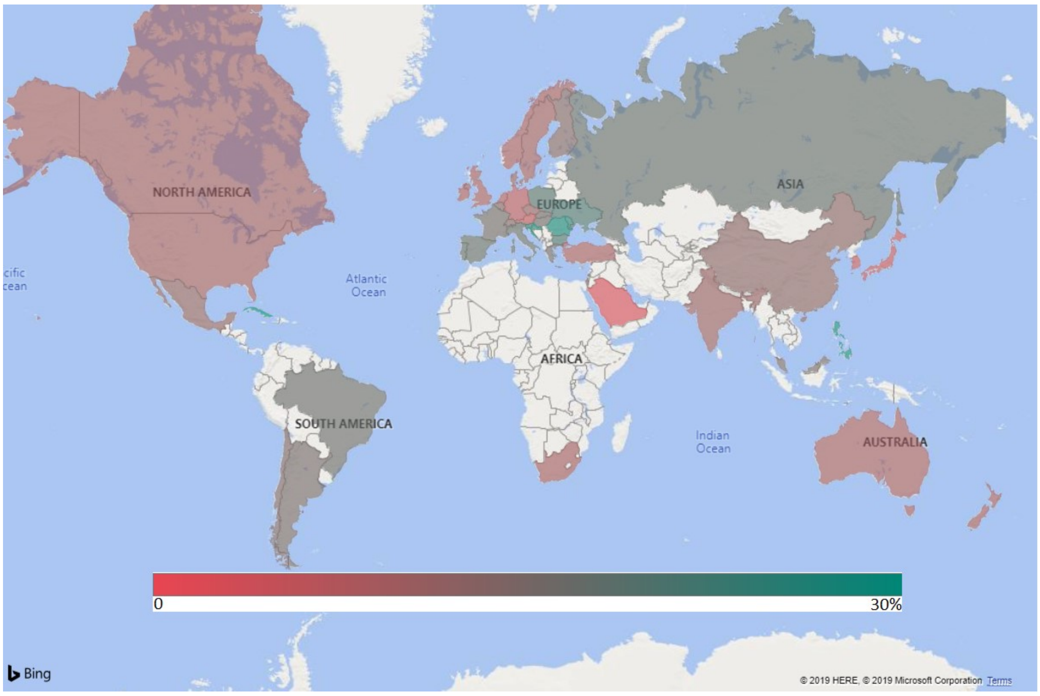


Figure 5: Proportion of female inventors for each of the top 50 countries, 1998-2017

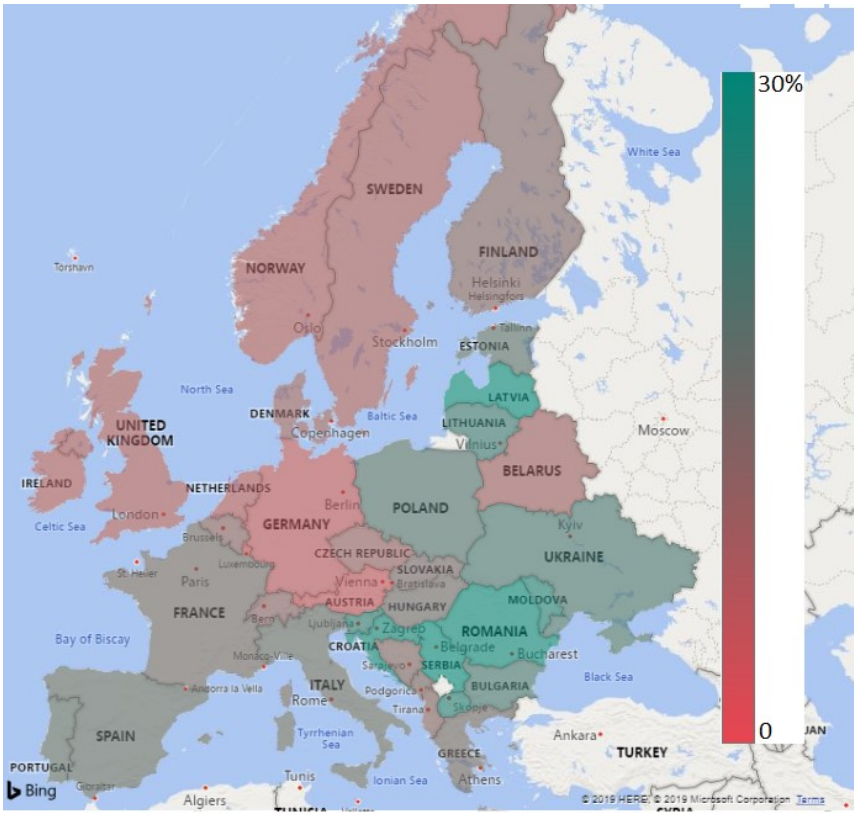


Figure 6: Proportion of female inventors for each country in Europe, 1998-2017

Figure 7 is a choropleth map that focusses on the United Kingdom. The data is generated using address information that is included in the PATSTAT dataset from inventors who declare their residence country as the United Kingdom. The postcode is extracted from the address field and mapped to either England, Scotland, Wales or Northern Ireland. There are over 10 times as many inventors in England than in the other three countries combined, and England has therefore been further subdivided into counties. A different colour scale is used for this map than for Figure 5 and Figure 6 because the United Kingdom has a lower proportion of female inventors than some other countries.

Of the four countries, Northern Ireland has the highest proportion of female inventors (10.6%), followed by Scotland and England (8.7%), whereas Wales has 8.2% female inventors. Within England, there is a general trend for regions in the South-East to have high proportions of female inventorship. Oxfordshire has 13.2% female inventors, Hertfordshire has 12.5%, Cambridgeshire has 11.7% and Greater London has 10.7%. Of the Northern regions, those with the highest proportions of female inventors are Tyne and Wear (13.4% female inventors), Nottinghamshire (11.6%) and Cheshire (11.2%).

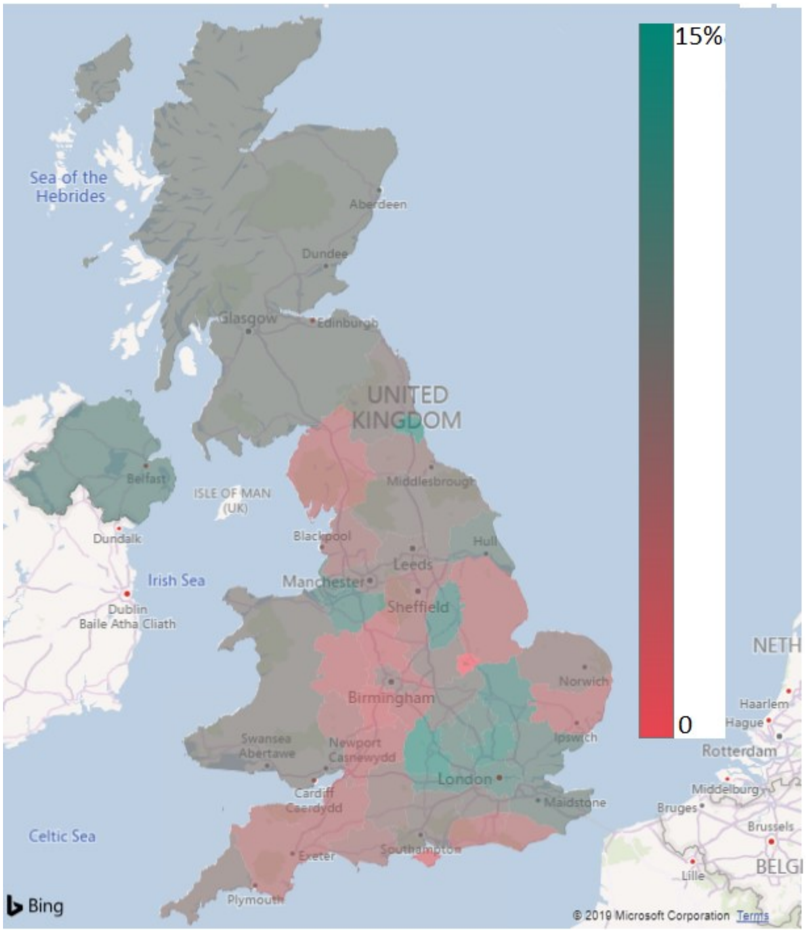


Figure 7: Proportion of female inventors in each region of the United Kingdom, 1998-2017

Female representation within teams of inventors

One area of particular interest when looking at inventor gender is how teams of inventors are made up of males and females. Each patent application can have one named inventor (a lone/individual inventor) or multiple inventors (working collaboratively as a team). By linking the inferred gender of each named inventor and the inventors listed on each patent application, we can analyse the extent to which female inventors work on their own, as part of a single-gender team, or as part of a mixed team.

Worldwide overview of inventor teams

Figure 8 relates to patents for which the inventors can be reliably inferred to be either an individual of known gender, an all-male team, an all-female team, or a mixed team[18]. Over the past 20 years, a growing number of these patents have had a female inventor on the team, rising from 13% in 1998 to 31% in 2017. The majority of female inventors participated as part of a team of mixed gender. In Figure 8, the dark blue line represents all mixed teams, whereas the pale blue line represents those mixed teams that have a lone female. This suggests that the majority of mixed teams of inventors have a lone female, and this trend appears to hold against a background trend of rising female participation in inventorship throughout the 20-year period shown. A similar conclusion can be reached when looking at the proportion of all-female teams that comprise only females, and those that comprise one female working individually. This suggests that females are much more likely to contribute to a patent application as a sole female inventor than jointly with other females.

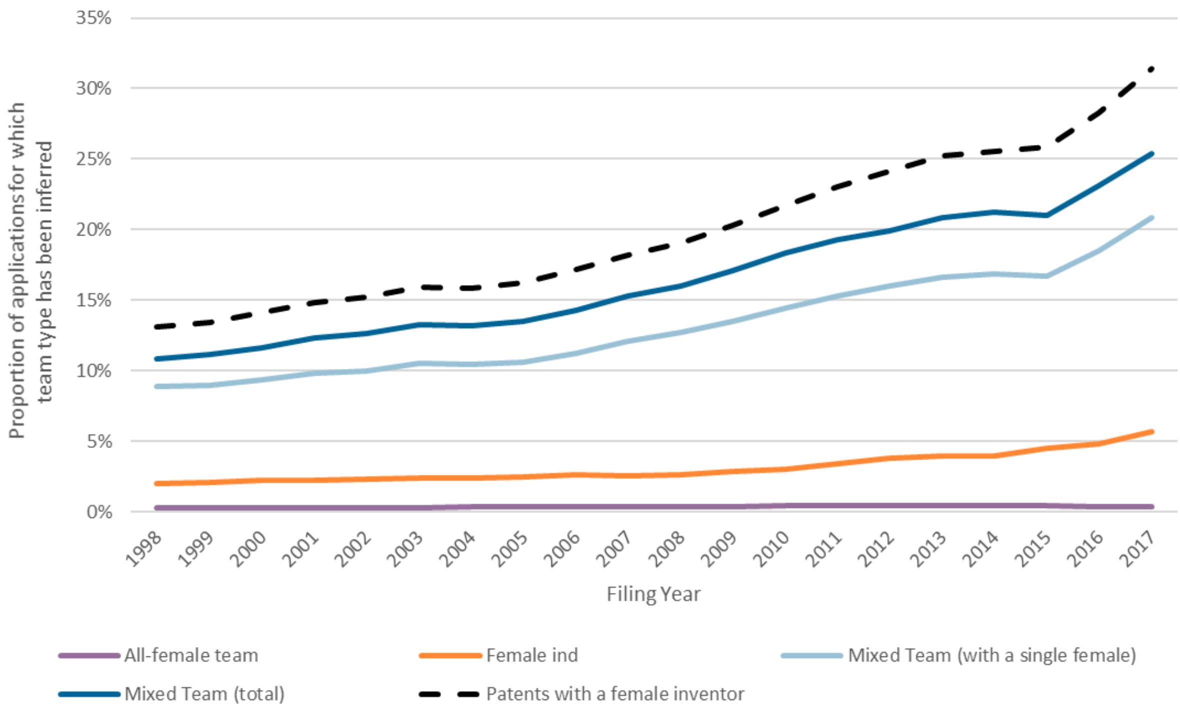


Figure 8: Proportions of patents associated with each category of team[19], 1998-2017

[18] Patent applications with, for example, 1 female inventor, 0 male inventor and 2 inventors of unknown gender have been omitted because it is impossible to determine whether the team is all-female or mixed. Due to the omission of patents for which the team type cannot be fully inferred, the quantitative proportions of patent applications shown in Figure 8 differs from Figure 3 (which focusses on inventors rather than patent applications), though the underlying trends are the same

[19] This figure only includes patent applications for which the team type can be fully inferred

Figure 9 compares the two five-year periods 1998-2002 and 2013-2017. Both time periods show a clear gender disparity, though there has nonetheless been a marked increase in the proportion of patent applications with at least one female inventor. As mentioned above, the large majority of patents for which all inventors were female had a lone inventor working alone, and this appears to have been the case both in 1998-2002 (2.2% patents with lone-female inventors out of 2.5% patents for which all inventors were female) and in 2013-2017 (4.5% lone-female inventors out of 4.9%). In contrast, of those patents for which all inventors are male, there are broadly similar numbers of applications with lone-male inventors as there are applications with all-male teams. This also appears to have been the case both in 1998-2002 (44.1% patents with lone-male inventors and 41.7% patents with all-male teams) and in 2013-2017 (40.5% patents with lone-male inventors and 32.2% all-male teams).

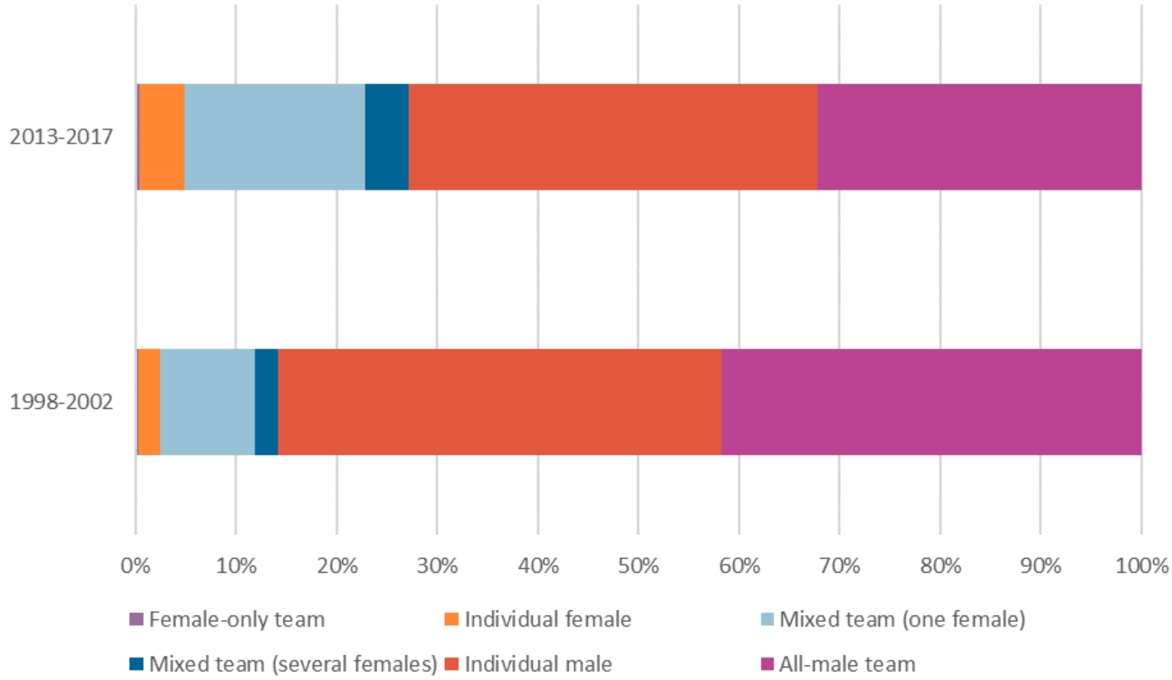


Figure 9: Comparing the inventor types[20] on patent applications between 1998-2002 and 2013-2017

[20] This figure only includes patent applications for which the team type can be fully inferred

Inventor team types for selected countries

Figure 10 shows how the inventor teams are distributed for the seven countries with the largest number of resident inventors[21]. Due to the differing inventor gender inference rates for each country, the gender inference rate for each country also varies. For ease of comparison, applications for which the team type cannot be fully inferred have been omitted, but the team coverage rate is listed alongside each country.

Figure 10 suggests that France, China, the USA and the UK all show relatively high proportions of female inventorship. France, and to a lesser extent the USA and the UK, are notable for having a larger proportion of patent applications with two or more female inventors. China and South Korea suggest a tendency for inventors to act as individuals rather than as part of a team, regardless of gender, but this could simply be an artefact of the lower coverage rate for those countries[22].

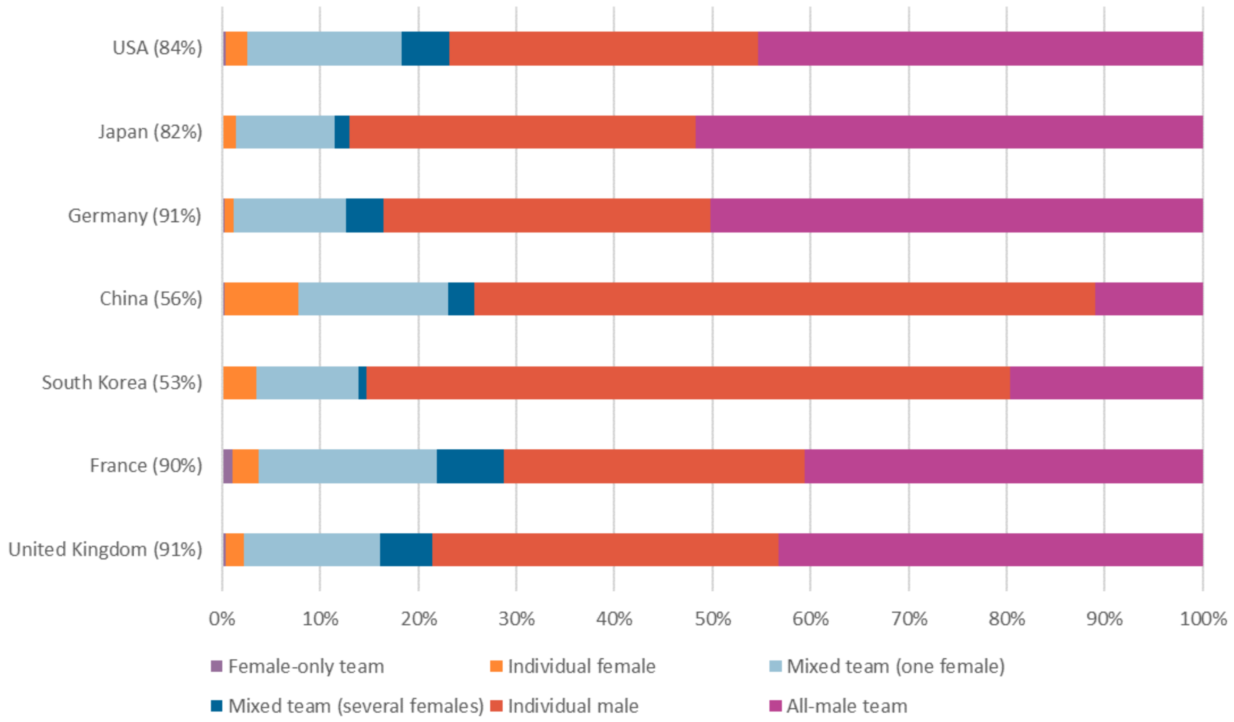


Figure 10: Inventor team constituency of patent applications[23] by residence country, 1998-2017 (Team constituency inference rates are given alongside each country)

[21] A patent application is counted for each country if there are one or more inventors resident in that country. It follows that a single patent may be counted multiple times if it has inventors that reside in different countries

[22] If the inventor gender inference rate is low, then it becomes increasingly likely for team gender information to be incomplete, and hence for the team to be omitted from the analysis, as the team size increases

[23] This figure only includes patent applications for which the team type can be fully inferred

The extent of female representation on mixed teams and on single-gender teams

Figure 11 and Figure 12 represent the extent to which female inventors participate in patent applications. An alternative way of quantifying female inventorship is by simply comparing the numbers of female and male inventors. These figures show that this may lead to different results.

Figure 11 shows the number of males on a mixed team with a lone female. Only 36% of these teams have a lone male, and the mean number of males is approximately 2.4. Hence, even though 17.9% of patent applications between 2013-2017 had mixed teams with a lone female, these applications would have had a somewhat larger number of male inventors than female inventors.

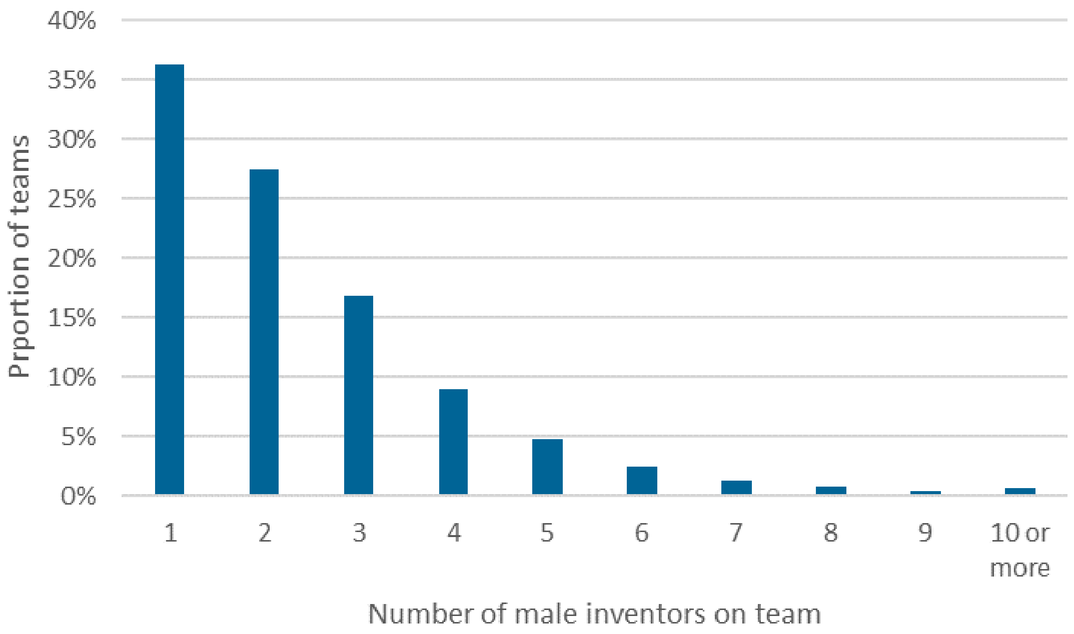


Figure 11: Distribution of the number of male inventors on inventor teams with a lone female inventor, 1998-2017

Figure 12 shows the number of inventors inferred as male (or female) on all-male (or all female) teams. A large majority (80%) of all-female teams are two-person teams, whereas only half of all-male teams are two person teams. Further, 10% of all-male teams have 5 or more inventors, whereas only 1% of all-female teams have so many inventors. The relatively large team sizes of all-male teams means that the proportion of inventors, across the entire dataset, that are male will be larger than may be inferred from Figure 8, Figure 9 and Figure 10. However, these figures nonetheless do indicate levels of female participation in patent inventorship, and Figure 8 in particular suggests that participation is increasing.

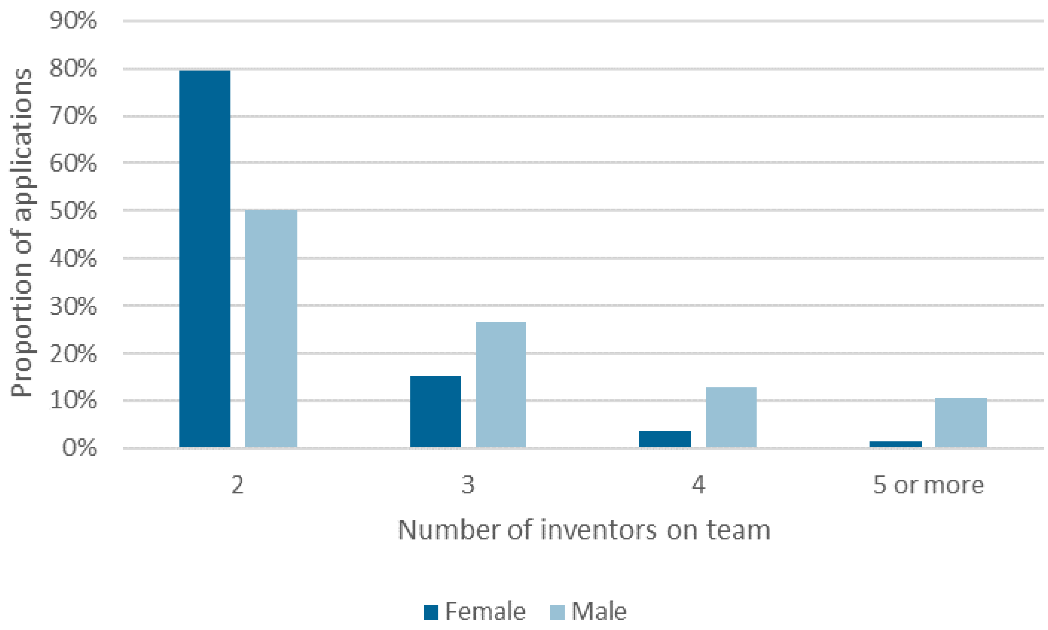


Figure 12: Number of inventors on teams that are either all-male or all-female, 1998-2017

Proportions of female inventorship within differing areas of technology

Patents may be sought for any idea with industrial application; patent applications therefore span many different technology subclasses. Applications are classified to one or more of these subclasses in accordance with the International Patent Classification (IPC) Scheme[24], and the classification of each patent application is available from the PATSTAT dataset[25]. There are over 650 different technology subclasses. In an attempt to gather together patents in similar technology areas, WIPO have published a concordance table[26], linking each subclass to one of 35 different technology "fields", which themselves fall under one of five broad technology "sectors". A patent application may represent developments in several technology subclasses, and so may be classified in several subclasses. In the analysis that follows, a patent is counted once against a field (or sector) if it is classified within one or more subclasses that fall within that field (or sector).

Technology sectors, fields and subclasses

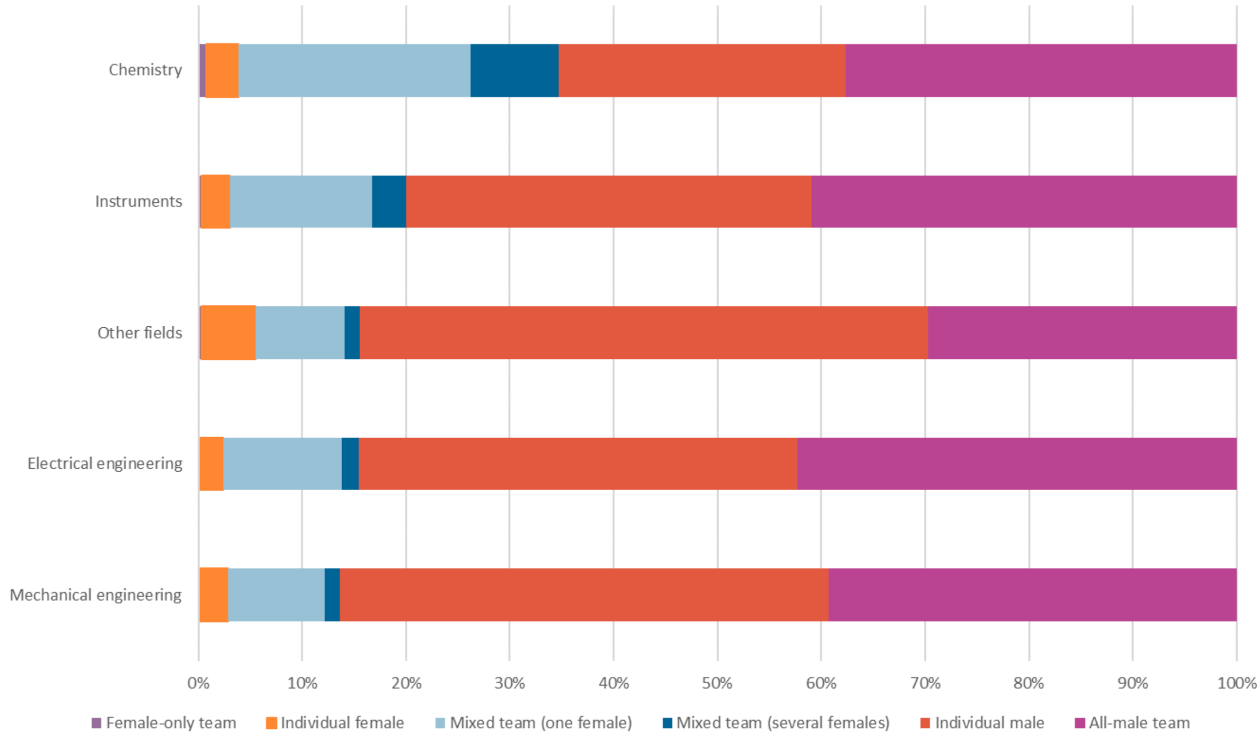


Figure 13: Inventor team constituency of patent applications by WIPO technology sector, 1998-2017

Figure 13 shows the proportions of inventor team types associated with patents within each technology sector. Although 65% of chemistry-related patents come from all-male inventors (28% male individuals and 38% all-male teams), female inventors have the biggest share in this sector with the remaining 35% of chemistry-related patents having at least one female inventor (3.1% female individuals, 0.8% all-female teams and 31%% mixed teams).

Apart from the "Chemistry" sector, the team types follow similar distributions for all technology fields. Again, it is notable that female participation in patent inventorship is typically of the form of sole inventorship or as a lone female in a mixed team. The "Other" sector, which covers miscellaneous subclasses not covered by the other four sectors, shows a relatively high proportion of sole female inventorship (5.2%), which is higher than

---

[24] Available at https://www.wipo.int/classifications/ipc/en/

[25] The relevant PATSTAT table is tls209_appln_ipc, and includes a list of application identifiers (appln_id) together with any IPC classification that has been applied (ipc_class_symbol)

[26] The WIPO IPC-Technology Concordance table is available at https://www.wipo.int/ipstats/en/index.html

the “Chemistry” sector (3.1%) despite the latter sector showing a higher proportion of patents with female inventors.

To analyse this in more detail, Figure 14 subdivides the five technology sectors into the 35 technology fields described above. The fields are arranged by sector and then ordered according to the proportion of patents with one or more female inventors.

Of the fields in the "Chemistry" technology sector, Biotechnology and Pharmaceuticals have the largest proportion of patents with female inventorship, with over half of the patents in these fields having a female inventor associated with them. A significant proportion of patents in these fields also have mixed teams comprising more than one female, in contrast to most of the other technology fields in Figure 14. Of the fields in the "Measurement" technology sector, the field with the largest proportion of patents with female inventorship is the analysis of biological materials (with 34%), followed by Medical technology and Measurement (both with 23%). Within the miscellaneous "Other" technology sector, which Figure 13 showed had a large proportion of patents with sole female inventors, the fields with the largest proportions of such patents are Furniture and games (with 6.1%) and other consumer goods (with 7.8%). The team breakdown for the field of Civil Engineering, which is also contained within the "Other" technology sector, is broadly in line with most of the fields within the "Electrical Engineering" and "Mechanical Engineering" sectors.

To provide yet further insight into how female inventorship varies between technology areas, Figure 15 shows the breakdown of teams within individual subclasses. The subclasses with the ten highest and ten lowest rates of female participation in patent inventorship are shown. The top ten subclasses all have female participation in over half of their patent applications, and all fall within either the “Biotechnology”, “Pharmaceuticals” and “Organic fine chemistry” technology fields. Conversely, the majority of the bottom ten subclasses fall within the “Mechanical Engineering” sector.

Figure 16 shows the IPC subclasses that have received the most applications (regardless of inventor gender) between 1998 and 2017. Amongst these subclasses are some of those in Figure 15 which have the highest levels of female participation in inventorship, which relate to pharmaceuticals and organic chemistry. This supports anecdotal evidence that these fields are areas of rapid technological development and enjoy high levels of female participation.

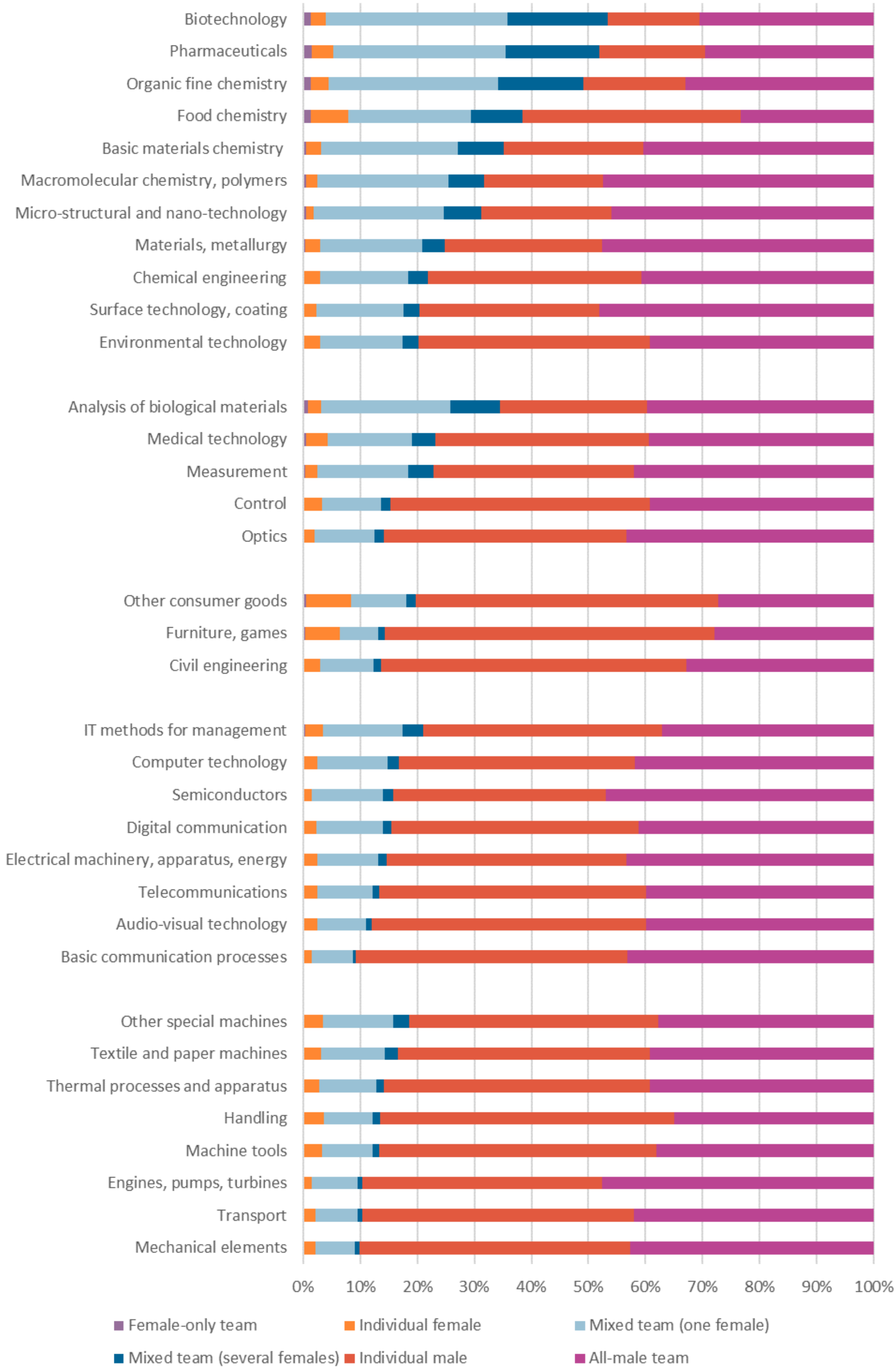


Figure 14: Inventor team constituency of patent applications by WIPO technology field, 1998-2017

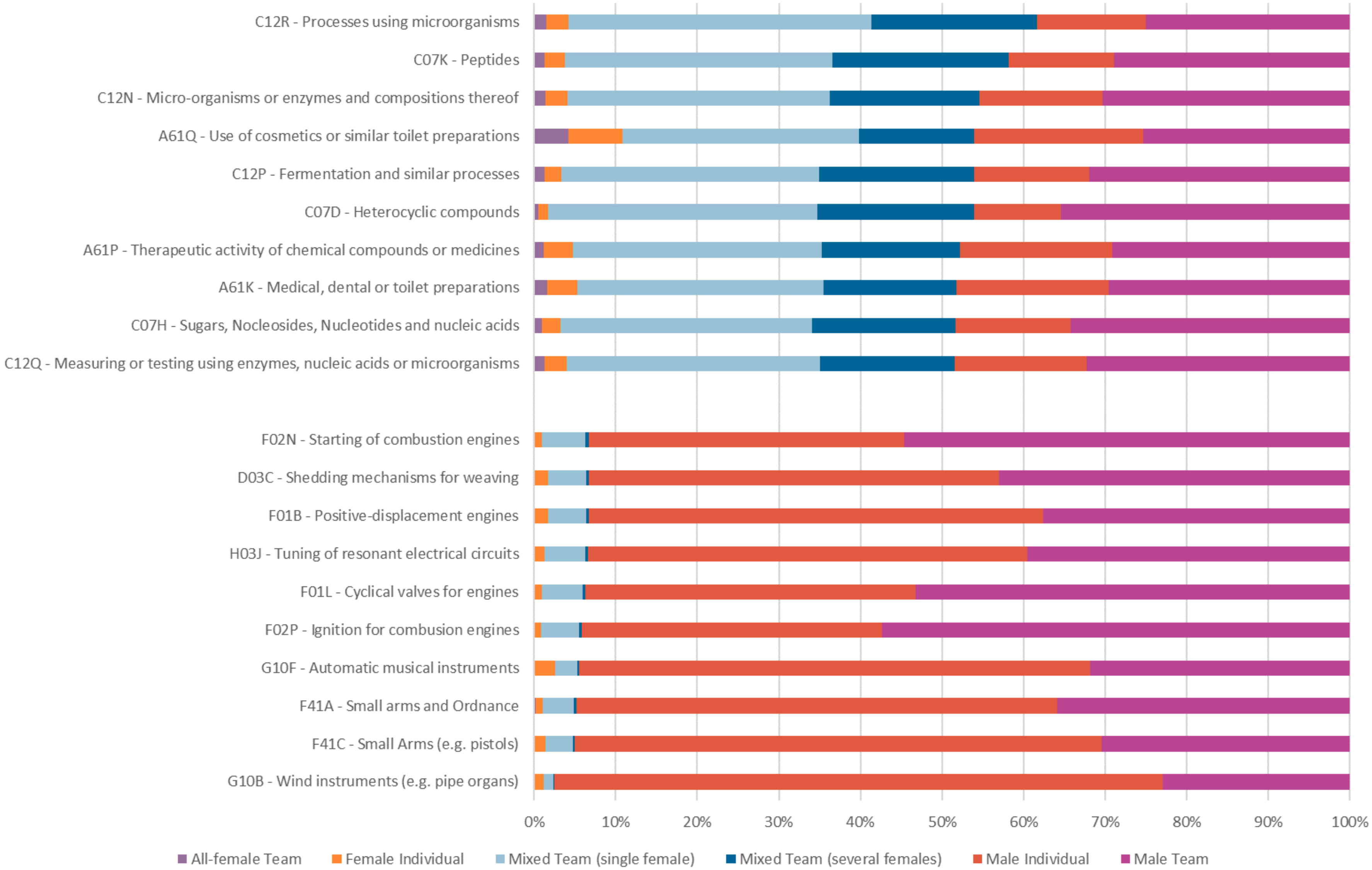


Figure 15: Inventor team constituency of patent applications by subclass (top 10 and bottom 10 by female participation), 1998-2017

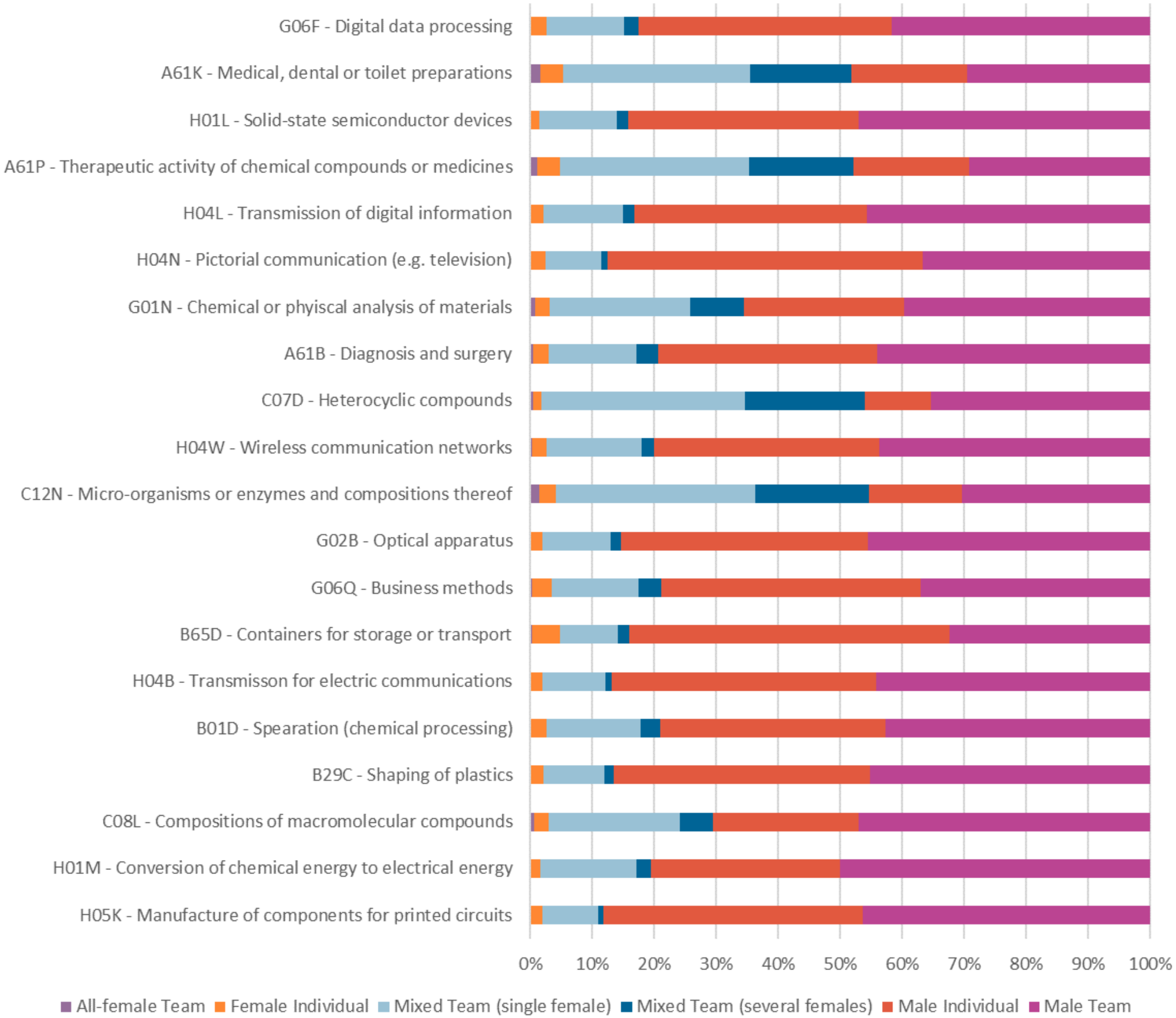


Figure 16: Inventor team constituency of patent applications by subclass (top 20 by application count), 1998-2017

Female and male inventorship across sections of the IPC scheme

Figure 17 shows the relative specialisation of males and females by the sections of the IPC. The height of each bar represents the proportion of each gender that has applied for one or more patents classified in that section. For each gender, the total height of the bars is greater than 100% due to some inventors working across several sections of the IPC, which provides some indication of the breadth of technical expertise of male and female inventors. On average, male and female inventors apply for patents within similar numbers of IPC sections, with males applying for patents classified in 1.64 sections and females applying for 1.60. This suggests that, although the absolute number of female inventors is much smaller than the number of male inventors, both genders have a similar breadth of expertise. There are also differences in the types of technology that are worked on by males and by females. For both genders, there is a low specialisation in sections D (Textiles and Paper), E (Fixed Constructions) and F (Mechanical Engineering, Lighting, Heating, Weapons and Blasting), which reflects a relatively low level of recent patenting activity in these sections. The highest proportion (33%) of males specialise in section G (Physics), though 28% of females also specialise in this section. The sections that show the largest gender disparity are section H (Electricity), in which 28% of males specialise but only 17% of females, and section A (Human necessities), in which 43% of females specialise but only 26% of males. Section C (Chemistry and Metallurgy) also show a large disparity, with 39% of females specialising in this section but only 25% of males.

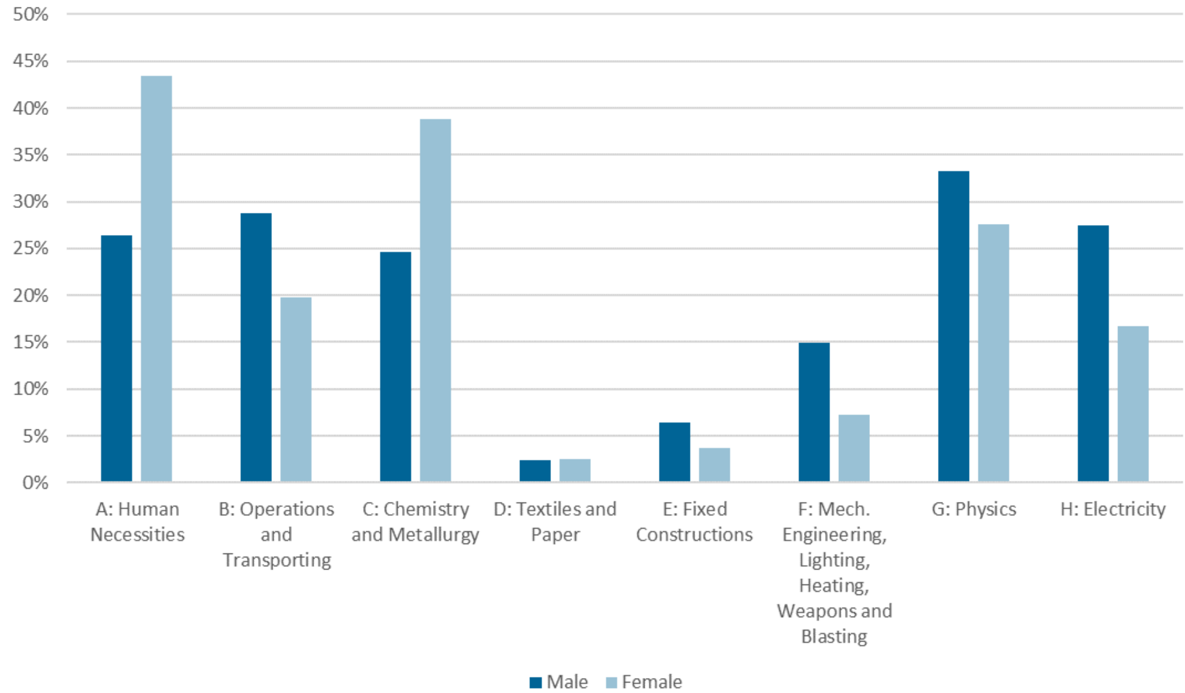


Figure 17: Relative specialisation of female and male inventors in each section of the IPC, 1998-2017

With the exception of sections D, E and F, similar numbers of male inventors specialise in each section of the IPC. In contrast, there are much larger numbers of female inventors specialising in sections A and C than the other sections. This is consistent with Figure 15, which shows that the 10 subclasses with the largest proportion of female participation are in sections A and C.

Proportion of female inventorship in academia or in industry

The PATSTAT database includes information on whether an identifier relates to (amongst other types) an individual, a company or a university. Not all applicant or inventor identifiers are attributed to a category, but those that are should give some indication as to how inventorship is split by gender in academia or by industry.

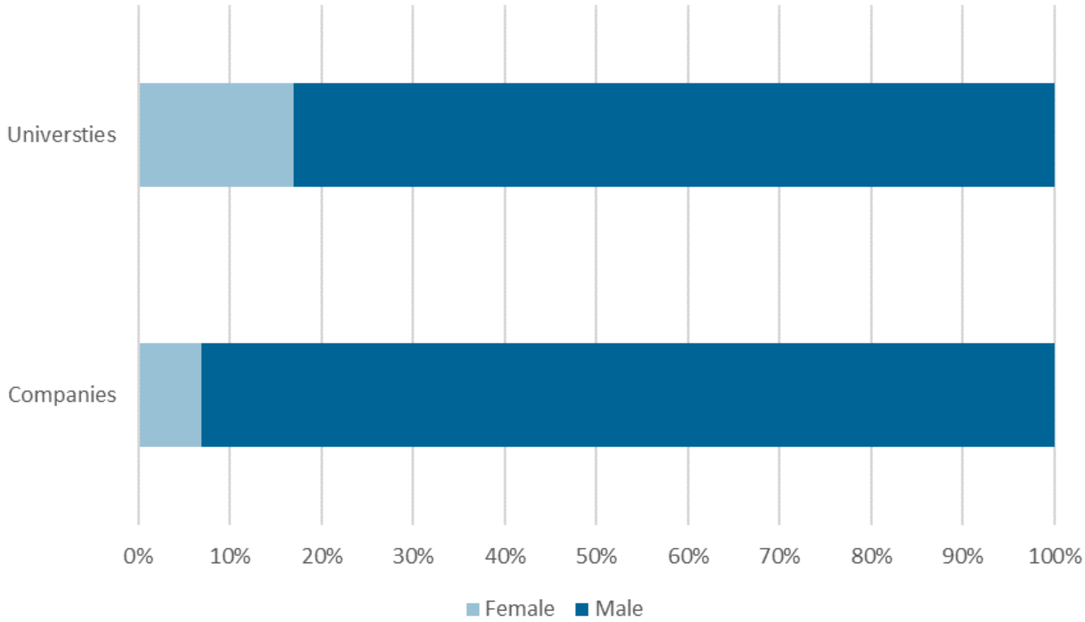


Figure 18: Proportions of female and male inventors associated with patent applications made by universities and by companies, 1998-2017

Figure 18 compares the number of female inventors associated with patents, filed between 1998-2017, for which an applicant is listed on PATSTAT as a company, and for which an applicant is listed as a university. There is a much higher proportion of female inventors associated with academia than with industry. This may reflect there being a differing environment in academia rather than industry. As Figure 18 illustrates, there are higher proportions of females undertaking higher education than go on to undertake careers in research. Additionally, inventions may be conceived during postdoctoral employment, during the course of a PhD, or perhaps even whilst studying for a Masters level degree. Such inventions may also contribute to the large proportion of female inventors exhibited in Figure 1.

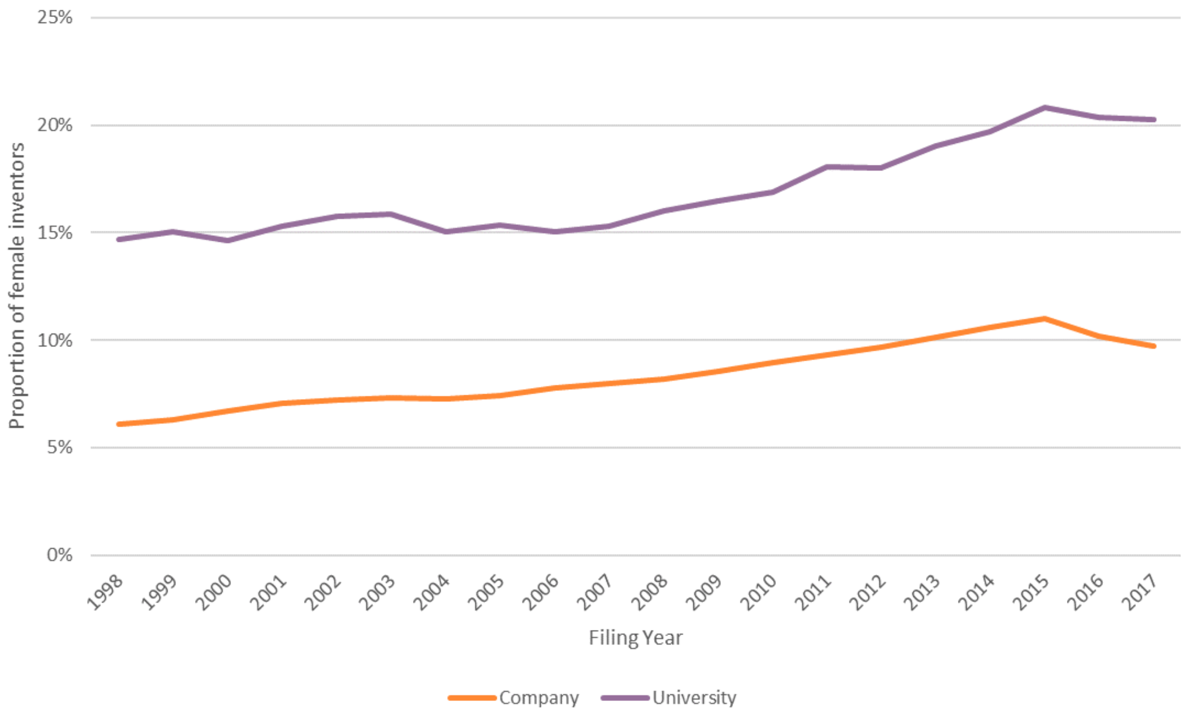


Figure 19: Time series of the proportions of female and male inventors associated with patent applications made by universities and by companies, 1998-2017

Figure 19 shows the year-on-year trends in proportions of female inventorship in academia and in industry. Although industry lags behind academia by around 15% throughout the period 1998-2017, it is reassuring that female inventorship is increasing in both sectors at comparable rates.

Discussion

This study suggests that many fewer inventors are female than are male, which is unsurprising in light of both anecdotal evidence that levels of female inventorship are low, and in light of other recent studies on female inventorship and on engagement with STEM education and careers. However, this study has shown that these levels are steadily increasing, in both industry and academia, and across the majority of countries. The results of this study also support anecdotal evidence that female inventorship is more common in academia than in industry.

Another perspective has been provided by considering the proportion of patents involving a female inventor. The majority of patents with a female inventor comprise only a lone female inventor, either as an individual inventor or as part of a mixed team that is typically male-dominated. Over the past 20 years, female participation has increased to the extent that over one in five patents now have at least one female inventor.

When broken down by technology type, it was shown that the fields of Biotechnology, Pharmaceuticals and Organic Chemistry enjoy the largest proportion of female participation, with females contributing to more than half of the first two of these fields. Several of these fields are amongst the most active technology area in terms of the number of applications filed (by an inventor of any gender).

As with 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 and attempted to improve coverage by using the PATSTAT dataset as a source of additional forenames and using an iterative approach to associate forenames with a gender. Although the iterative approach has improved coverage, the initial data sources[27] are inherently biased towards ‘western’ first names and so such an approach will inevitably be incomplete when inferring the gender of other inventors, such as those with Asian names. In future studies, wider-ranging datasets could be sought to improve coverage.

The analysis contained in this report provides an overview of the type of patent analysis that can be undertaken using the inferred inventor gender table generated by the IPO of the UK, which has been published alongside this report[28]. 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*[29].

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.

[27] See the Technical Annex

[28] Available at https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship-2019-edition

[29] Available at https://www.gov.uk/government/publications/the-patent-guide

Technical annex

Data sources

Patent data

This study, like the earlier studies undertaken by the IPO of the UK[30,31], attempts to infer inventor gender using data from the European Patent Office (EPO) Worldwide Patent Statistical Database (PATSTAT)[32]. This dataset is widely used by patent researchers and academics worldwide and so was the natural choice for this study. It contains bibliographic data relating to almost 100 million patent documents from leading industrialised and developing countries[33] and provides a comprehensive collection of data which can be set up and queried using Structured Query Language (SQL). The IPO of the UK is keen to make the outputs of this study widely available and so a table has been published alongside this report[34] that associates the inventor identifiers used by PATSTAT[35] with their inferred gender, where found, using the methodology described below.

Initial name-gender dataset

Some recent academic research projects looking at inferring gender from name data have made it possible to infer the gender of inventors on patent applications. This study has relied on the results of two different methodologies, one originating from Massachusetts Institute of Technology (MIT)[36] (hereinafter known as the Matias methodology) and the other from Peking University/NYU Polytechnic School of Engineering/Max Planck Institute for Software Systems[37] (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 IPO of the UK 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 purposes of the present study the IPO of the UK used only 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

---

30 Gender profiles in UK patenting: An analysis of female inventorship, April 2016. Available at https://www.gov.uk/government/publications/gender-profiles-in-uk-patenting-an-analysis-of-female-inventorship

31 Gender profiles in worldwide patenting: An analysis of female inventorship, November 2016. Available at https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship

32 The PATSTAT version used in the present study is the 2019 Spring Edition (2019a)

33 https://www.epo.org/searching-for-patents/business/patstat.html

34 Available at https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship-2019-edition

35 Each person_id is associated with a single inventor in PATSTAT, and so is used in our table as a key.

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

37 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

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. The IPO of the UK 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. In order to provide a high-quality dataset for analysis, only names for which at least 95% of entries are male or female are inferred to be the respective gender. Names, such as “Robin”, which are commonly used both by males and by females, are therefore not inferred as either gender due to insufficient confidence.

In order to provide a high-quality dataset for analysis, the confidence threshold for gender inference was set at 95%. A combination name-gender dataset consisting of data from both the Matias and Tang methodologies was used as a seed dataset for this study. In a manner similar to that undertaken by WIPO[38], we used an iterative methodology to infer the gender of forenames that could not be inferred directly from the combination name-gender dataset. This is described in further detail in the Methodology subsection of this Annex.

Initial surname dataset

The PATSTAT dataset is generated using data submitted by patent offices across the world. Consequently, a person might have their name recorded as “Joe Bloggs”, “Bloggs, Joe”, or “Bloggs Joe” depending on the patent office that submitted the data. To assist in extracting forenames correctly, an initial dataset of surnames based on US census data was used[39].

[38] Martinez et al., “Identifying the gender of PCT inventors”. WIPO Economic Research Working Paper No. 33, November 2016. Available at https://www.wipo.int/publications/en/details.jsp?id=4125

[39] The dataset is available at: https://www.census.gov/topics/population/genealogy/data/2010_surnames.html

Methodology

Data extraction

Bibliographic information relating to the patent applications contained in PATSTAT[40] 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[41]. Sequence numbering is used to separate person identifiers relating to applicants and inventors[42] and thus it is possible, at least in principle, to extract a list of all person identifiers with sequence numbering that shows that they are inventors and not applicants (*e.g.* companies).

The 2019 Spring Edition of PATSTAT contains 49,619,209 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.

Data cleaning and data linking

The inventor name data held in PATSTAT is generated using information supplied to the EPO from the individual patent registers of national and regional patent offices around the world. Unfortunately, the name data has no standard format and so some preliminary cleaning is necessary. 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.

The gender inference rate of the previous study of the IPO of the UK was 73.4% of all inventor identifiers for which a forename match was feasible (corresponding to 95.8% of all identifiers). This inference rate was obtained by first extracting the first name for each inventor and then inferring the gender based on that name. In the present study, coverage is extended using an iterative approach that relies on some inventors having multiple forenames. For example, an inventor named "Jane Josephina Doe" would be inferred to be female based on the forename "Jane". If the name "Josephina" were absent from the name-gender table, it could be inferred as a female name if the majority of inventors with that forename are inferred to be female. On this basis, "Josephina" could be added to the name-gender table. This process may then be repeated iteratively to further improve inference rates to around 75%. However, the success of this process depends on accurate extraction of inventor forenames.

Because name data in PATSTAT has no standard format, a preliminary obstacle is to attempt to infer the name format used by each jurisdiction (which may also vary from time to time). The approach taken was to generate an initial dictionary that attributes words to one of the following categories:

- Forename – any name of three of more letters that is captured in the combined Mathias-Tang dataset
- Surname – any name appearing in the US Census dataset
- Forename or Surname – any name appearing both "Forename" and "Surname" categories
- Title – Honorifics such as Dr, Prof, Dipl
- Suffices – for example Jnr, Esq, III, deceased
- Initials – single-letter words
- Company – for example "Corp", "PLC", "Technology"
- Unknown – any name not included in the above.

[40] https://www.epo.org/searching-for-patents/business/patstat.html
[41] In PATSTAT the applicant identifier is appln_id and the person identifier is person_id
[42] In PATSTAT the inventor sequence number (invt_seq_nr) is 0 for applicants and 1 or more for inventors

This dictionary was then used to attribute formats to each name, by first splitting it into words and punctuation marks, such as in the following:

- “Bloggs, Joe J.” → “U,FI.”
- “Prof. Doe, Jane” → “T.U,I”

Upon inspection of the data it was found that information about the company was sometimes included in the person_name field (e.g. “Joe Bloggs, Acme Corp.”), and sometimes the person_name field consisted entirely of a company name. Because many companies are named after people, it is important to remove as many company names as feasible before attempting to infer the gender of the inventor. This was done by assuming that the inventor’s name would appear before that of the company, and then using the following steps:

- Identify a name containing a word associated with a company rather than a person
- Read back from that word until either a punctuation mark or “c/o” is found
- Delete everything from that punctuation mark onwards or, if there is no such punctuation mark, remove the entire string

The intended effect is that names like “Joe Bloggs, Acme Corp.” will be converted to “Joe Bloggs”, and names like “Acme Corp.” will be deleted entirely.

The deleted words were checked manually to extract additional words relating to companies. Simply relying on the frequency of deleted words proved to be unreliable – although one might expect that generic words such as “Corp.” would appear much more frequently than company-specific words like “Acme”, several company names occur with high frequency due to the sheer number of patent applications they file.

It was also found that the person_name field often contained numerical information relating to addresses. Another cleaning step was therefore taken where any word containing digits, and any subsequent words, were deleted from the name. Finally, words relating to either titles or suffices, and anything contained in brackets, were removed from the string.

After the above cleaning steps are completed, each word in the cleaned person_name field is highly likely to be either a surname, forename, initial, or of “unknown” type. To infer the gender of an inventor, their forenames need to be identified, but this is made difficult by differing name formats used by different patent offices. Generally, the name format either has the surname appearing first, as in “Bloggs, Joe Joseph” or last, as in “Joe Joseph Bloggs”. The forename and surname datasets were therefore used to determine whether each patent office used the “surname first” or “surname last” name format. Because the name format may vary with time, the name format was determined on a per-office per-year basis.

Until this point in the data cleaning, the attribution of words as surnames and forenames depends entirely on the coverage of the Matias-Tang (for forenames) and US Census (for surnames) datasets. To extend this coverage, an iterative approach was used where countries which use a consistent name format were used to infer additional surnames and forenames. This was done by classifying each name format as having either the surname located first, last, or at an unknown position in the cleaned name. Then, a patent office was determined to use a consistent format in a given year if either of the following conditions held:

- There are 9 times as many names for which the surname appears first compared to the number of names for which the surname appears last (or vice versa)
- The surname appears first 10 or more times and there are no occurrences of the surname appearing last (or vice versa)

For a given patent office that employs a consistent name format in a given year, all inventors associated with that office and that year were inferred to follow the same format, and this inference was relied upon to extract additional forenames and surnames. The process was then repeated, using the additional forenames and surnames, to infer information about applications from other patent offices.

Having inferred the name format used by the majority of patent offices each year, forenames were extracted from the PATSTAT dataset as follows:

1. If the name format is known for the patent office and the year of the application, assign forenames based on that format
2. Otherwise:
    a. If there is one surname in the name, assign all other words as forenames
    b. If there are multiple surnames, then discard the name on grounds of ambiguity (as it potentially includes multiple inventors)
    c. If all words are "forename or surname", discard the name on grounds of ambiguity (as the name "James Stewart" could equally be "Stewart James")

Step 2 was carried out twice, the first time to extract additional forenames and surnames using sub-step a), and the second time to infer additional person names using a dictionary that was expanded during the first iteration.

Name genders were then inferred initially using the name-gender concordance table that was obtained by the 2016 study[31] as explained above. In the present study, the forename extraction algorithm used can yield multiple forenames per inventor, thereby increasing the likelihood of assigning a name to a gender. Ambiguous names, for which a "male" forename and a "female" forename are present for the same inventor, were excluded from the study. Inventors with one or more "male" forenames and no "female" forenames were inferred to be male, and vice versa. It is tempting to extrapolate these inferences to provide information about the gender associated with forenames of as-yet-unknown gender. However, care is needed due to there being many more male inventors than female inventors in the PATSTAT dataset.

The approach taken in the present study is to infer that a name represents a male or female if it is associated with the respective gender 95% of the time. However, there are roughly 10 times as many male inventors as female inventors in the PATSTAT dataset, which has the effect of overstating the number of males associated with a given name and understating the number of females. If 95% of people (both inventors and non-inventors) called "Chris" are male, then it can be shown that on average, 99.5% of people called "Chris" in the PATSTAT dataset would be male. Conversely, if 95% of people called "Jules" are female, then on average 66.5% of people called "Jules" in the PATSTAT dataset would be female. The thresholds of 99.5% and 66.5% are therefore used to infer additional male forenames and female forenames.

After inferring additional forename genders, inventors with previously-uninferred gender can now be processed. The determination of additional forename genders, and the inference of additional inventor genders, was repeated 5 times.

Research output

Accompanying this report, the IPO of the UK has published[43] a table which can be used in conjunction with PATSTAT containing the inventor gender disaggregation for 33,326,210 identifiers in PATSTAT. Also provided is a spreadsheet containing the data shown in the figures.

[43] Available at https://www.gov.uk/government/publications/gender-profiles-in-worldwide-patenting-an-analysis-of-female-inventorship-2019-edition

Name-gender dataset overview

An overview of the final dataset used for this study is provided in Figure 20. It shows that 33.3 million of the 44.4 million 'matchable' (i.e. with a forename) unique inventor identifiers (75.1%) in PATSTAT (*i.e.* inventors for which one or more forenames were extracted) have successfully had a gender inferred with a 95% certainty. This equates to 67.2% of all unique inventor identifiers in PATSTAT (33.3 million out of 49.6 million unique inventor identifiers). These inventor identifiers are associated with 55.6 million patent applications (75.7% of all applications that have inventor information available on PATSTAT) and 42.1 million of those applications have the gender inferred for all inventors. It may well be possible to improve coverage by reducing the confidence threshold below 95%, but this has not been pursued in this study.

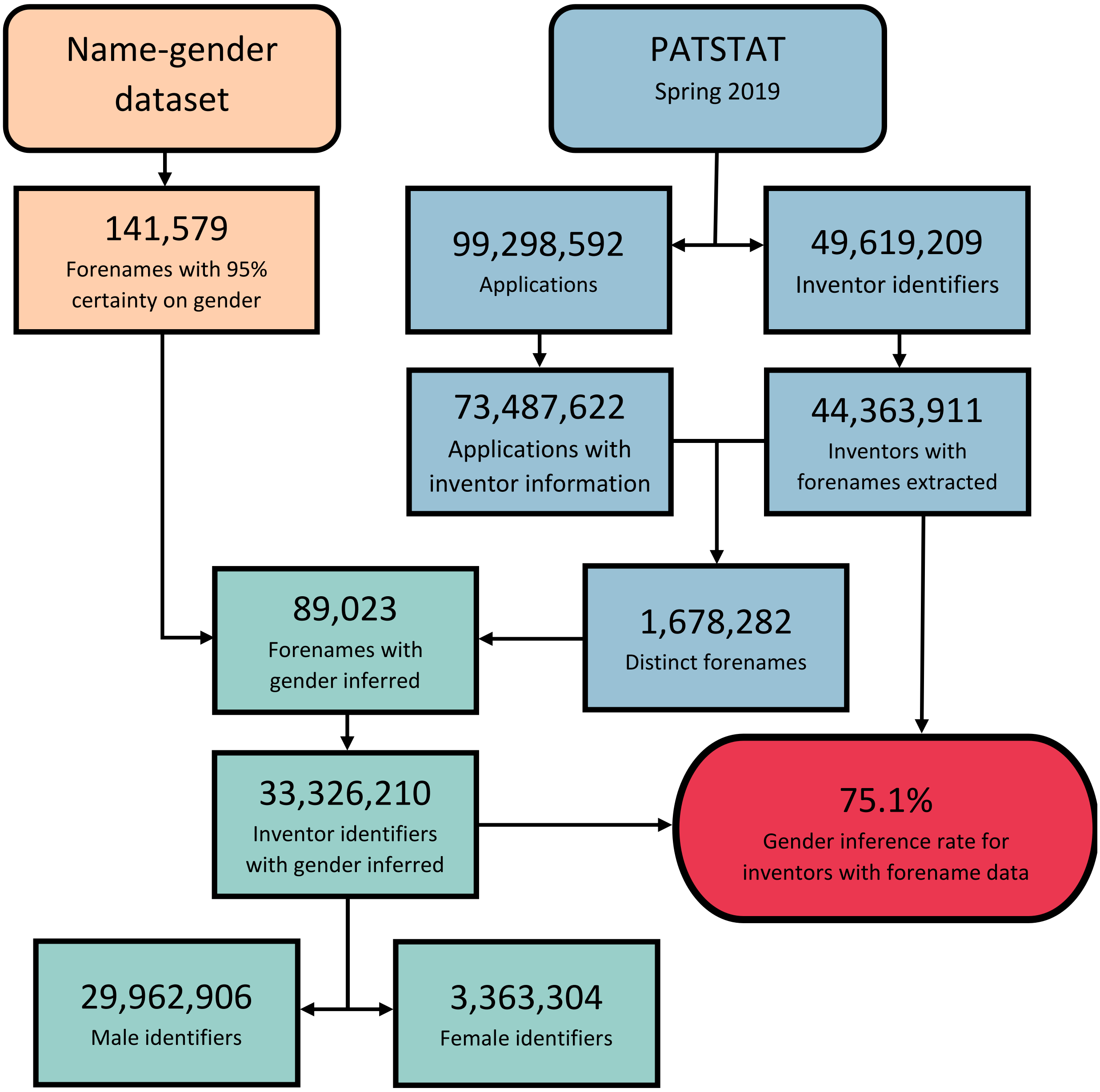


Figure 20: Dataset overview

Dataset coverage

Figure 21 shows the inference rate according to the declared country of residence, for the 15 countries with the largest numbers of resident inventors. Because forenames are more likely to correlate with where a person resides than where a person files their patent applications, it was felt more appropriate to analyse coverage on the basis of residence country rather than the jurisdiction of the patent. Although PATSTAT has only incomplete inventor residency information, there is enough coverage to provide a representative sample.

For the countries shown, inference rates are generally very good at around 90%. The exceptions to this are Taiwan, China and South Korea. This highlights a deficiency in our name-gender dataset in that it is biased towards Western names. Although the iterative approach used in the present study would have captured some non-Western names, there will be many forenames that do not appear often enough for a reliable inference to be made about the inventor's gender. The easiest way to improve coverage in future work would be to seek additional sources of forenames to expand the name-gender dataset to cover more non-Western names.

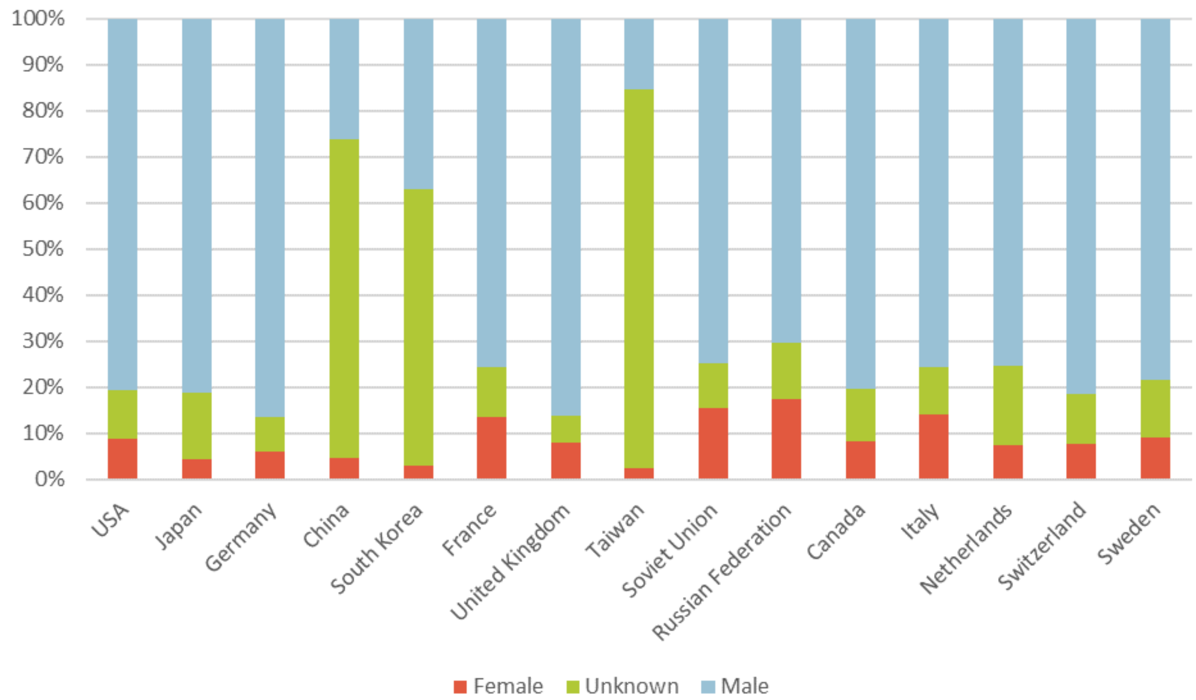


Figure 21: Gender inference rates by declared country of residence

Figure 22 plots a breakdown of the gender breakdown of all patents filed between 1998 and 2017 for which at least one inventor was inferred to be male or female (patents for which no inventor had an inferred gender are thus omitted). Of these patents, approximately 26% could not be fully inferred as they had, for example, two female inventors, no male inventors and one inventor of uninferred gender. Because these applications may correspond either to mixed teams or to single-gender teams, they are omitted from the commentary in the report.

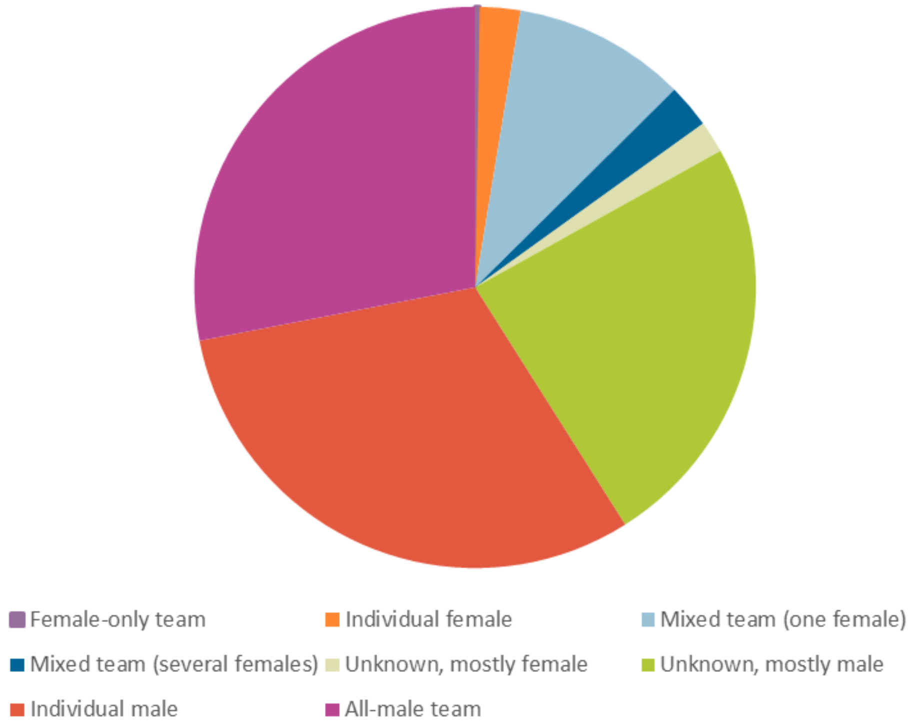


Figure 22: Proportions of patents associated with each category of team, 1998-2017

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
LinkedIn: uk-ipo

For copies in alternative formats please contact our Information Centre.

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

Published: October 2019