Immigration, Gender, and Doctoral Dreariness

Aramis D. M. Valverde

12/10/2021

Is It Perhaps Different Doing A PhD In A Country You Grew Up In As Opposed to One That You Moved To?

I Mean Probably, IDK Though, Lets Find Out!

Abstract: A 2019 Nature survey of doctoral students worldwide conducted by Shift Learning, a UK based education market research and consulting firm, demonstrated that PhD students are not doing all that well. The Nature write up by Chris Woolston noted that more than a third of PhD students have sought help for anxiety or depression caused by their their PhD studies, and more than a fifth of PhD students had experienced discrimination or harassment in their PhD program. Prior research has backed up the concept that PhD students, and graduate students at large, aren't doing terribly well. As a PhD student I thought, "yeah, I feel it, its pretty bad, I'm pretty sad.". However I reassured myself that at least I didn't have to deal with being an actual immigrant or a woman. But then one asks oneself, "huh, is it actually worse?. Maybe Bill O'Rielly was right when he said that women don't feel discrimination in the workplace."

To analyze if it is worse to be a woman or a gender queer/ non binary person, or an immigrant, as opposed to not those things, as a PhD student, I analyzed Nature's data set as published on figshare.

The results were xyz, abc, def, ghi, jkl, mno, pqr, stu, vwx, yza.

https://group.springernature.com/gp/group/media/press-releases/archive-2019/nature-phd-survey-puts-spotlight-on-mental-health/17372858 https://www.zmescience.com/other/pieces/journals-to-blame-poor-phd-mental-health-0432/ https://www.insidehighered.com/news/2019/11/14/phd-student-poll-finds-mental-health-bullying-and-career-uncertainty-are-top https://www.nature.com/articles/d41586-019-03535-y https://figshare.com/articles/dataset/Data_publication_survey_raw_data/1234052 https://figshare.com/articles/dataset/Nature_Graduate_Survey_2017/5480716?file=9558301 https://www.nature.com/articles/nj7677-549a

https://phys.org/news/2019-11-nature-phd-survey-spotlight-mental.html https://www.nature.com/articles/d41586-021-01538-2 https://www.apollo.io/people/Karen/Kaplan/54a7df517468696b7f4ceb54 https://laptrinhx.com/news/interview-with-karen-kaplan-senior-careers-editor-at-nature-E3OYqY5/

Introduction: Things are hard for PhD students. Research has demonstrated a persistent and severe toll on the mental health of PhD students due to the conditions that they are put under. More than a third have sought help for anxiety and depression caused by their PhD studies, and a fifth of PhD students have experienced discrimination and harassment in their program. A 2020 study of UK PhD students found that 40 percent of said students were at high risk of suicide, with 8 percent having made an attempt.

That being said, every doctoral student's experience is not the same. Some student's experiences are worse. U.S. based doctoral students originating from China also have to deal with isolation, loneliness, and alienation. Some trans students have to deal with severe hostility or insensitivity related to their gender identity or expression, leading suicide attempts rates four times higher than trans students who did not have that severe experience. It is therefore very unfortunate that this survey lumped transgendered and cisgendered

people into one variable. Erasure, especially in research, has consequential and damaging effects on trans people.Gender queer and non-conforming was included, so I will attempt an analysis. However people identifying as such were few, so I may not be able to demonstrate a significant difference even if one is liable to exist. It is also important to note that gender-specific approaches yield more accurate results, contribute to lessening the paucity of trans research, and are considered best practice.

I must note that I am a minority doctoral student myself. While I am a minority in my department and in my country, I am not an immigrant, nor am I a minority in the state or city where I am studying (Merced, California). I struggle with no language barrier, no one chides me for my voice, and I can travel freely throughout my campus and the city I reside in. I note this because, while my analysis of the data stands on its own, my identity is such that I am not able to speak from direct experience, and that is a limitation within this work.

I was lead to this dataset by a tidytuesday post on github, and this research was completed as the final assignment for Dr. Dan Hicks' data science methods course at UC Merced. The data set used in this analysis was sourced from this figshare page.

Methods

The PhD Student Survey's History Nature has had a survey of doctoral students conducted every two years for ten years. The latest survey, conducted in 2019, was advertised by Nature in its "Career News" section, and was titled, "Nature calling: take our PhD survey". This survey was conducted by an English corporation by the name of Shift Insight. Shift Insight is the owning and operating body of Shift Learning, Shift Sustainability, and Shift Membership https://shift-insight.co.uk/about-shift-insight/. This corporation is owned by Jane Powellhttps://www.linkedin.com/in/jenny-kedros-3074471a/, a former Publisher at Pearson and Palgrave Macmillan. Shift Insight mainly performs market researchhttps://shift-insight.co.uk/research-reports-and-blogs/, however as with the Nature survey they also conduct public facing research for large institutions. Most staff at Shift Insights have a marketing, writing, or publishing background.

Their website's language seems to demonstrate commitments to data securityhttps://shift-insight.co.uk/quality-and-values/, however these amount to either declarations of compliance strictly within the confines of the law, or membership with non-binding commerce and industry associations. Unfortunately for participants, the 2019 survey failed to remove all personally identifying information, specifically the emails of two students and the names of two others, according to my own analysis where I found two names and emails.

The finance director and data compliance officer, Jack Wilsonhttps://www.linkedin.com/in/jack-wilson-0751853/, appears to have no credentials to handle said data, unlike some of his co-workers, and seems to have been educated at "LAMBDA" without any named degree, according to their linkedin profile. While I could not locate a university by the name of LAMBDA, it is possible that this education was associated with a fraternal organization,and it is also quite possible that Jack Wilson graduated with that education at the top of their class. I mention this not because I want to dunk on these people, but rather because their seemingly irresponsible handling of data could have some real consequences for those people who have had their data exposed. I also mention this because Shift appears to have existed for 20 years but has only had moderately to mildly credentialed persons on staff for seven yearshttps://www.linkedin.com/search/results/people/?currentCompany=%5B%222294535%22%5D&origin=COMPANY_I with the only visible doctorate holder entering the company November 2021https://www.linkedin.com/in/daisy-alero-emoekabu-phd-b22b2319/. This, along with the founders' recent certification and education in R may help explain why the data arrived in a state that was not in line with tidy data principles, and why the analysis of the data was relatively limited. The privacy statement, limited as it was, may have also been broken by the lack of proper anonymizationhttps://shift-insight.co.uk/privacy-policy/.

https://shift-insight.co.uk/about-shift-insight/#our-team

The PhD Student Survey's Line of Custody The data appears to have passed through the hands of at least three people, Elsie Lauchlanhttps://www.linkedin.com/in/elsie-lauchlan-990276a3/,

the primary researcher for the analysis, David Paynehttps://www.linkedin.com/in/david-payne-0a681013/?originalSubdomain=uk, The managing editor at Careers and Supplements at Nature Springer, and Karen Kaplanhttps://www.linkedin.com/in/karen-kaplan-03186b9/, the senior editor at Nature's careers section. According to the version control sheet in the survey script, there were four versions of the analysis or survey. That is very few, and further analysis could have been forsaken due to cost, as the same page notes that additional changes may involve extra costs.

The complete list of questions is available in the data documentation, specifically the .docx file therein.

Limitations The survey has a couple of issues with it. First is the fact that the data has been cleaned to an extent that has not been disclosed by shift insights. The amount of anonymization and the methods of cleaning the data from the vendor or software that was used to produce the data is not fully and clearly disclosed. Because of this we have no way of discerning what was removed or how the data was manipulated. I also could not attempt the ordinal surevy, as the link had expired., https://wh1.snapsurveys.com/s.asp?k=162210914288&s=article it does appear to have been conducted through Snapsurveys.comhttps://www.snapsurveys.com/survey-software/pricing/.

Second is the fact that the participant pool is liable to have been skewed because of the method of advertisement (via Nature Careers) and because Shift Insights seems to reuse participantshttps://shift-insight.co.uk/respondent-faqs/. Even within this survey there were questions and language to indicate that this survey was also used to recruit to their participant pool. Because of their commercial interests and focus on market research, it is possible that their existing participant pool is incidentally selecting against some traits. For example, I am not liable to respond to an email from some random corporation doing market research, because I don't respect market researchers. This is partially mitigated against by the fact that some PhD students appear to have thought that Nature was conducting the survey and not Shift Insights, as they responded as if the people conducting the survey had some sort of authority in spite of the fact that they've nothing of the sort.

Third is the fact that the survey was only conducted in the aforementioned five languages. Persons from other parts of the world who communicate in languages that arent English, Portugese, iohu;aerg aeruhkljn and hioaerg; are not likely to have responded to the survey. There is also the issue that persons from certain countries were not given the option to select their country. Taiwan made up the vast majority of write in answers for country of origin in Asia.

Fourth is a lack of representation. There is little mention or analysis of non cisgendered persons, and there is no capacity for me to take a gender specific analysis of the data as mentioned prior.

Finally, we have the issue that these are self reported data and that that participants may have wanted to introduce their own positive self perceptions. Some participants may have had some concerns about putting this information in a website for fear that their comments may make them identifiable. For at least four students, these sorts of concerns would have certainly been warranted. The participants may have also wanted to present themselves in a positive light more generally. Given the doctoral community is more than a little neurotic, there are a myriad of manners by which this data could have been skewed.

Code and Analysis paste this in to hide output, code, and code: {r someVar,results='hide', echo=FALSE,include=FALSE}

Load in Libraries and Installations

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.2
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.5      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.0      v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
## Warning: package 'tidyr' was built under R version 4.1.2
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

```
## Warning: package 'forcats' was built under R version 4.1.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.1.2
```

```
library(ggplot2)
library(tinytex)
```

```
## Warning: package 'tinytex' was built under R version 4.1.2
```

```
library(RColorBrewer)
library(ggdist)
```

```
## Warning: package 'ggdist' was built under R version 4.1.2
```

```
library(waffle)
```

```
## Warning: package 'waffle' was built under R version 4.1.2
```

```
library(dplyr)
```

Load in data set that I downloaded from the link above(and which I placed into a "data" folder within the project folder) and check that it imported properly

```
dfog <- read_excel("data/Nature_PhD survey_Anon_v1.xlsx")
```

This is meant to ensure that it all was imported properly

```
glimpse(dfog)
```

```
## Rows: 6,813
## Columns: 274
## $ ID.format    <chr> "The published format which was employed", "SNAP 2015 SHI~
## $ ID.completed <chr> "Case completed in Snap Interviewer", "completed", "compl~
## $ ID.language  <chr> "What language would you like to complete the survey in?"~
## $ ID.site      <chr> "Questionnaire location", NA, NA, NA, NA, NA, NA, NA, NA,~
## $ ID.date      <chr> "Date of interview", "06/14/2019", "06/14/2019", "06/14/2~
## $ ID.start     <chr> "Time interview started", "0.69908564814814811", "0.71530~
## $ ID.endDate   <chr> "Completion date of interview", "06/14/2019", "06/14/2019~
## $ ID.end       <chr> "Time interview ended", "0.73346064814814815", "0.7350115~
## $ ID.time      <chr> "Duration of interview", "49.5", "28.38", "23.87", "55.5"~
## $ Q1           <chr> "Which, if any, of the following degrees are you currentl~
## $ Q1.a         <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q2           <chr> "Hidden", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Q3           <chr> "Which was the most important reason you decided to enrol~
## $ Q3.a         <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q4           <chr> "Are you studying in the country you grew up in?  ", "Yes~
## $ Q5           <chr> "Where do you currently live?", "North or Central America~
## $ Q6           <chr> "Which region in Asia?", NA, NA, NA, NA, NA, NA, "India",~
## $ Q6.a         <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q7           <chr> "Which country in Australasia?", NA, NA, NA, NA, NA, NA, ~
## $ Q7.a         <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q8           <chr> "Which country in Africa?", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q8.a         <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q9           <chr> "Which country in Europe?", NA, NA, NA, NA, NA, NA, NA, "~
## $ Q9.a         <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q10          <chr> "Which country in North or Central America?", "Mexico", "~
## $ Q10.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q11          <chr> "Which country in South America?", NA, NA, NA, NA, NA, NA~
## $ Q11.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ `Q12:1`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:2`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:3`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:4`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:5`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:6`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:7`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:8`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:9`      <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:10`     <chr> "What prompted you to study outside your country of upbri~
## $ `Q12:11`     <chr> "What prompted you to study outside your country of upbri~
## $ Q12.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q13          <chr> "Do you have a job alongside your studies?", "No", "No", ~
## $ Q14          <chr> "What is your main reason for having a job?", NA, NA, NA,~
## $ Q14.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q15.a        <chr> "The difficulty of getting funding / low success rates fo~
## $ Q15.b        <chr> "Inability to finish my studies in the time period I had ~
## $ Q15.c        <chr> "Impact of a poor relationship with my supervisor/PI", NA~
## $ Q15.d        <chr> "The number of available faculty research jobs beyond pos~
## $ Q15.e        <chr> "The high numbers of PhD holders who are doing or have do~
## $ Q15.f        <chr> "The difficulty of maintaining a work/life balance", "3rd~
## $ Q15.g        <chr> "Uncertainty about the value of a PhD", "4th", "2nd", NA,~
## $ Q15.h        <chr> "Uncertainty about my job/career prospects", NA, "1st", N~
## $ Q15.i        <chr> "Student debt during my PhD", NA, NA, NA, NA, "1st", "14t~
```

```
## $ Q15.j        <chr> "Financial worries after my PhD (cost of living, inabilit~
## $ Q15.k        <chr> "Political landscape", NA, "3rd", NA, NA, NA, "12th", "6t~
## $ Q15.l        <chr> "Impostor syndrome", "2nd", NA, NA, NA, "6th", "10th", "8~
## $ Q15.m        <chr> "Concern about my mental health as a result of PhD study"~
## $ Q15.n        <chr> "Poor support and acknowledgement of my parenting/elder c~
## $ Q16          <chr> "Is there anything else not mentioned that has concerned ~
## $ Q17          <chr> "Overall, what do you enjoy most about life as a PhD stud~
## $ Q17.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q18.a        <chr> "How satisfied are you with your decision to pursue a PhD~
## $ Q19.a        <chr> "How satisfied are you with your PhD experience?", "5", "~
## $ Q20          <chr> "Since the very start of your graduate school experience,~
## $ Q21.a        <chr> "Availability of funding", "6", "7 = Extremely satisfied"~
## $ Q21.b        <chr> "Hours worked", "1 = Not at all satisfied", "7 = Extremel~
## $ Q21.c        <chr> "Social environment", "2", "5", "6", "5", "4 = Neither sa~
## $ Q21.d        <chr> "Degree of independence", "6", "7 = Extremely satisfied",~
## $ Q21.e        <chr> "Recognition from supervisor/PI", "4 = Neither satisfied ~
## $ Q21.f        <chr> "Overall relationship with supervisor/PI", "5", "7 = Extr~
## $ Q21.g        <chr> "Opportunities to collaborate", "6", "6", "6", "4 = Neith~
## $ Q21.h        <chr> "Number of publications", "1 = Not at all satisfied", "7 ~
## $ Q21.i        <chr> "Stipend / financial support", "6", "5", "3", "2", "1 = N~
## $ Q22.a        <chr> "Vacation time", "1 = Not at all satisfied", "3", "6", "3~
## $ Q22.b        <chr> "Benefits (health care, leave, etc.)", "1 = Not at all sa~
## $ Q22.c        <chr> "Teaching duties", "4 = Neither satisfied nor dissatisfie~
## $ Q22.d        <chr> "Guidance received from adviser in lab/research", "5", "5~
## $ Q22.e        <chr> "Guidance received from other mentors in lab/research", "~
## $ Q22.f        <chr> "Ability to attend meetings and conferences", "3", "7 = E~
## $ Q22.g        <chr> "Ability to present research at conferences", "3", "7 = E~
## $ Q22.h        <chr> "Work-life balance", "1 = Not at all satisfied", "6", "4 ~
## $ Q22.i        <chr> "Career pathway guidance and advice", "3", "5", "5", "2",~
## $ Q23          <chr> "To what extent does your PhD programme compare to your o~
## $ Q24          <chr> "On average, how many hours a week do you typically spend~
## $ Q25          <chr> "On average, how much one-on-one contact time do you spen~
## $ Q25.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q26          <chr> "Overall, how would you describe the academic system, bas~
## $ Q27.a        <chr> "Members of my department make time for frank conversatio~
## $ Q27.b        <chr> "Members of my department are open to the idea of me purs~
## $ Q27.c        <chr> "Members of my department have useful advice for careers ~
## $ Q27.d        <chr> "Members of my department have contacted potential employ~
## $ Q27.e        <chr> "Members of my department have encouraged me to attend ca~
## $ Q27.f        <chr> "Members of my department have discouraged me from attend~
## $ Q28          <chr> "Have you ever sought help for anxiety or depression caus~
## $ Q29          <chr> "Did you seek help for anxiety or depression within your ~
## $ Q29.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q30.a        <chr> "Mental health and wellbeing services in my university ar~
## $ Q30.b        <chr> "My supervisor/PI has a good awareness of support service~
## $ Q30.c        <chr> "My university offers adequate one-to-one mental health s~
## $ Q30.d        <chr> "My university offers  different types of support to prom~
## $ Q30.e        <chr> "My university supports good work-life balance", "Strongl~
## $ Q30.f        <chr> "There is a long-hours culture at my university, includin~
## $ Q31          <chr> "Do you feel that you have experienced bullying in your P~
## $ `Q32:1`      <chr> "Who was the perpetrator(s)?", NA, NA, NA, NA, NA, NA, NA~
## $ `Q32:2`      <chr> "Who was the perpetrator(s)?", NA, NA, NA, NA, NA, NA, NA~
## $ `Q32:3`      <chr> "Who was the perpetrator(s)?", NA, NA, NA, NA, NA, NA, NA~
## $ `Q32:4`      <chr> "Who was the perpetrator(s)?", NA, NA, NA, NA, NA, NA, NA~
```

```
## $ `Q32:5`      <chr> "Who was the perpetrator(s)?", NA, NA, NA, NA, NA, NA, NA~
## $ `Q32:6`      <chr> "Who was the perpetrator(s)?", NA, NA, NA, NA, NA, NA, NA~
## $ `Q32:7`      <chr> "Who was the perpetrator(s)?", NA, NA, NA, NA, NA, NA, NA~
## $ Q32.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q33          <chr> "Do you feel able to speak out about your experiences of ~
## $ Q34          <chr> "Do you feel that you have experienceddiscrimination or h~
## $ `Q35:1`      <chr> "Which of the following have you experienced?", NA, "Raci~
## $ `Q35:2`      <chr> "Which of the following have you experienced?", NA, NA, N~
## $ `Q35:3`      <chr> "Which of the following have you experienced?", NA, NA, N~
## $ `Q35:4`      <chr> "Which of the following have you experienced?", NA, NA, N~
## $ `Q35:5`      <chr> "Which of the following have you experienced?", NA, NA, N~
## $ `Q35:6`      <chr> "Which of the following have you experienced?", NA, NA, N~
## $ `Q35:7`      <chr> "Which of the following have you experienced?", NA, NA, N~
## $ `Q35:8`      <chr> "Which of the following have you experienced?", "Other, p~
## $ `Q35:9`      <chr> "Which of the following have you experienced?", NA, NA, N~
## $ Q35.a        <chr> "If other, please specify", "Por pertenecer a un programa~
## $ Q36.a        <chr> NA, "Unsure", "Somewhat", "Dramatically", "Substantially"~
## $ Q37.a        <chr> "Academia", "4th", "1st", "1st", NA, "5th", "2nd", "1st",~
## $ Q37.b        <chr> "Industry", "1st", "5th", "4th", "2nd", "4th", "4th", "3r~
## $ Q37.c        <chr> "Government", "5th", "2nd", "2nd", NA, "3rd", "3rd", "2nd~
## $ Q37.d        <chr> "Non-profit", "2nd", "3rd", NA, "1st", "2nd", "5th", "5th~
## $ Q37.e        <chr> "Medical", "3rd", "4th", "3rd", NA, "1st", "1st", "4th", ~
## $ Q38.a        <chr> "Research in academia", "Not very likely", "Very Likely",~
## $ Q38.b        <chr> "Research in industry", "Neither likely nor unlikely", "L~
## $ Q38.c        <chr> "Research within government or non-profit", "Not very lik~
## $ Q38.d        <chr> "Non-research in academia", "Not likely at all", "Likely"~
## $ Q38.e        <chr> "Medical research", "Likely", "Unsure", "Likely", "Not ve~
## $ Q38.f        <chr> "Non-research in industry", "Very Likely", "Unsure", "Not~
## $ Q38.g        <chr> "Non-research in government or non-profit", "Not likely a~
## $ `Q39:1`      <chr> "If you're unlikely to pursue an academic research career~
## $ `Q39:2`      <chr> "If you're unlikely to pursue an academic research career~
## $ `Q39:3`      <chr> "If you're unlikely to pursue an academic research career~
## $ `Q39:4`      <chr> "If you're unlikely to pursue an academic research career~
## $ `Q39:5`      <chr> "If you're unlikely to pursue an academic research career~
## $ `Q39:6`      <chr> "If you're unlikely to pursue an academic research career~
## $ `Q39:7`      <chr> "If you're unlikely to pursue an academic research career~
## $ `Q39:8`      <chr> "If you're unlikely to pursue an academic research career~
## $ `Q39:9`      <chr> "If you're unlikely to pursue an academic research career~
## $ Q39.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q40          <chr> "What position do you most expect to occupy immediately a~
## $ Q40.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q41          <chr> "What type of career you are interested in pursuing after~
## $ Q42          <chr> "After completing your PhD, how long do you think it will~
## $ Q43.a        <chr> "How much more likely are you now to pursue a research ca~
## $ Q44          <chr> "What is the main reason why you are more likely to pursu~
## $ Q44.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ `Q45:1`      <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:2`      <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:3`      <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:4`      <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:5`      <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:6`      <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:7`      <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:8`      <chr> "How did you arrive at your current career decision?  Ple~
```

```
## $ `Q45:9`    <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:10`   <chr> "How did you arrive at your current career decision?  Ple~
## $ `Q45:11`   <chr> "How did you arrive at your current career decision?  Ple~
## $ Q45.a      <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ `Q46:1`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:2`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:3`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:4`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:5`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:6`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:7`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:8`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:9`    <chr> "How do you learn about available career opportunities th~
## $ `Q46:10`   <chr> "How do you learn about available career opportunities th~
## $ `Q46:11`   <chr> "How do you learn about available career opportunities th~
## $ `Q46:12`   <chr> "How do you learn about available career opportunities th~
## $ `Q46:13`   <chr> "How do you learn about available career opportunities th~
## $ `Q46:14`   <chr> "How do you learn about available career opportunities th~
## $ Q46.a      <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ `Q47:1`    <chr> "Which of the following 3 things would you say are the mo~
## $ `Q47:2`    <chr> "Which of the following 3 things would you say are the mo~
## $ `Q47:3`    <chr> "Which of the following 3 things would you say are the mo~
## $ `Q47:4`    <chr> "Which of the following 3 things would you say are the mo~
## $ `Q47:5`    <chr> "Which of the following 3 things would you say are the mo~
## $ `Q47:6`    <chr> "Which of the following 3 things would you say are the mo~
## $ `Q47:7`    <chr> "Which of the following 3 things would you say are the mo~
## $ `Q47:8`    <chr> "Which of the following 3 things would you say are the mo~
## $ `Q47:9`    <chr> "Which of the following 3 things would you say are the mo~
## $ Q47.a      <chr> "If other, please specify", NA, "Finding career opportuni~
## $ `Q48:1`    <chr> "Which of the following would you say are the most diffic~
## $ `Q48:2`    <chr> "Which of the following would you say are the most diffic~
## $ `Q48:3`    <chr> "Which of the following would you say are the most diffic~
## $ `Q48:4`    <chr> "Which of the following would you say are the most diffic~
## $ `Q48:5`    <chr> "Which of the following would you say are the most diffic~
## $ `Q48:6`    <chr> "Which of the following would you say are the most diffic~
## $ `Q48:7`    <chr> "Which of the following would you say are the most diffic~
## $ Q48.a      <chr> "If other, please specify", NA, "Finding career opportuni~
## $ `Q49:1`    <chr> "Which of the following resources do you think PhD studen~
## $ `Q49:2`    <chr> "Which of the following resources do you think PhD studen~
## $ `Q49:3`    <chr> "Which of the following resources do you think PhD studen~
## $ `Q49:4`    <chr> "Which of the following resources do you think PhD studen~
## $ `Q49:5`    <chr> "Which of the following resources do you think PhD studen~
## $ `Q49:6`    <chr> "Which of the following resources do you think PhD studen~
## $ `Q49:7`    <chr> "Which of the following resources do you think PhD studen~
## $ `Q49:8`    <chr> "Which of the following resources do you think PhD studen~
## $ Q49.a      <chr> "If other, please specify", NA, "merit based selections r~
## $ Q50.a      <chr> "Collecting data", "Well", "Well", "Very well", "Well", "~
## $ Q50.b      <chr> "Analysing data", "Well", "Well", "Well", "Well", "Very w~
## $ Q50.c      <chr> "Designing robust reproducible experiments", "Well", "Nei~
## $ Q50.d      <chr> "Writing a paper for publication in a peer-reviewed journ~
## $ Q50.e      <chr> "Developing  resilience to manage rejection by a peer rev~
## $ Q50.f      <chr> "Presenting findings to a specialist audience", "Well", "~
## $ Q50.g      <chr> "Presenting findings to a non-specialist (public) audienc~
## $ Q50.h      <chr> "Applying for funding", "Neither well nor badly", "Neithe~
```

```
## $ Q50.i        <chr> "Finding a satisfying career", "Badly", "Badly", "Well", ~
## $ Q50.j        <chr> "Managing complex projects", "Badly", "Badly", "Well", "W~
## $ Q50.k        <chr> "Developing a business plan", "Neither well nor badly", "~
## $ Q50.l        <chr> "Managing people", "Very badly", "Badly", "Badly", "Well"~
## $ Q50.m        <chr> "Managing a large operational budget", "Very badly", "Bad~
## $ Q51.a        <chr> "I feel that my programme is preparing me well for a rese~
## $ Q51.b        <chr> "I feel that my programme is preparing me well for a non-~
## $ Q51.c        <chr> "I feel that my programme is preparing me well for a care~
## $ ‘Q52:1‘      <chr> "Which, if any, of the following activities have you done~
## $ ‘Q52:2‘      <chr> "Which, if any, of the following activities have you done~
## $ ‘Q52:3‘      <chr> "Which, if any, of the following activities have you done~
## $ ‘Q52:4‘      <chr> "Which, if any, of the following activities have you done~
## $ ‘Q52:5‘      <chr> "Which, if any, of the following activities have you done~
## $ ‘Q52:6‘      <chr> "Which, if any, of the following activities have you done~
## $ ‘Q52:7‘      <chr> "Which, if any, of the following activities have you done~
## $ ‘Q52:8‘      <chr> "Which, if any, of the following activities have you done~
## $ Q52.a        <chr> "If other, please specify", "Escuché tutoriales acerca de~
## $ ‘Q53:1‘      <chr> "Which of the following social media networks have you us~
## $ ‘Q53:2‘      <chr> "Which of the following social media networks have you us~
## $ ‘Q53:3‘      <chr> "Which of the following social media networks have you us~
## $ ‘Q53:4‘      <chr> "Which of the following social media networks have you us~
## $ ‘Q53:5‘      <chr> "Which of the following social media networks have you us~
## $ ‘Q53:6‘      <chr> "Which of the following social media networks have you us~
## $ ‘Q53:7‘      <chr> "Which of the following social media networks have you us~
## $ Q53.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ ‘Q54:1‘      <chr> "What would you do differently right now if you were star~
## $ ‘Q54:2‘      <chr> "What would you do differently right now if you were star~
## $ ‘Q54:3‘      <chr> "What would you do differently right now if you were star~
## $ ‘Q54:4‘      <chr> "What would you do differently right now if you were star~
## $ ‘Q54:5‘      <chr> "What would you do differently right now if you were star~
## $ Q54.a        <chr> "If other, please specify", "Trataría de organizar mejor ~
## $ Q55          <chr> "With the benefit of hindsight, what one thing do you kno~
## $ Q56          <chr> "What is your age?", "25 - 34", "25 - 34", "25 - 34", "25~
## $ Q57          <chr> "Are you...", "Female (including trans female)", "Male (inc~
## $ ‘Q58:1‘      <chr> "Which of the following best describes you?", NA, NA, "Ca~
## $ ‘Q58:2‘      <chr> "Which of the following best describes you?", "Latino/His~
## $ ‘Q58:3‘      <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ ‘Q58:4‘      <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ ‘Q58:5‘      <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ ‘Q58:6‘      <chr> "Which of the following best describes you?", NA, "South ~
## $ ‘Q58:7‘      <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ ‘Q58:8‘      <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ ‘Q58:9‘      <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ ‘Q58:10‘     <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ ‘Q58:11‘     <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ ‘Q58:12‘     <chr> "Which of the following best describes you?", NA, NA, NA,~
## $ Q58.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ ‘Q59:1‘      <chr> "Do you have any caring responsibilities?", NA, NA, NA, N~
## $ ‘Q59:2‘      <chr> "Do you have any caring responsibilities?", NA, NA, NA, N~
## $ ‘Q59:3‘      <chr> "Do you have any caring responsibilities?", "Yes, to an a~
## $ ‘Q59:4‘      <chr> "Do you have any caring responsibilities?", NA, "No", "No~
## $ ‘Q59:5‘      <chr> "Do you have any caring responsibilities?", NA, NA, NA, N~
## $ Q59.a        <chr> "If other, please specify", NA, NA, NA, NA, NA, NA, NA, N~
## $ Q60          <chr> "Thank you for taking part in the survey. Are there any m~
```

```
## $ Q61          <chr> "Would you like to be entered into the prize draw to win ~
## $ Q62          <chr> "Nature may want to contact you again to ask for more inf~
## $ Q63          <chr> "Springer Nature is keen to update PhD students with advi~
## $ Q64          <chr> "Shift Learning carry out paid research in the education ~
## $ Q65.a        <chr> "Name:", "1", "2", "3", "4", "5", "6", "7", "8", "9", "10~
## $ Q65.b        <chr> "Email address:", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

Just to check that the top line has been preserved.

```
head(dfog)
```

```
## # A tibble: 6 x 274
##   ID.format  ID.completed ID.language ID.site ID.date ID.start ID.endDate ID.end
##   <chr>      <chr>        <chr>       <chr>   <chr>   <chr>    <chr>      <chr>
## 1 The publi~ Case comple~ What langu~ Questi~ Date o~ Time in~ Completio~ Time ~
## 2 SNAP 2015~ completed    Spanish     <NA>    06/14/~ 0.69908~ 06/14/2019 0.733~
## 3 SNAP 2015~ completed    English     <NA>    06/14/~ 0.71530~ 06/14/2019 0.735~
## 4 SNAP 2015~ completed    English     <NA>    06/14/~ 0.71899~ 06/14/2019 0.735~
## 5 SNAP 2015~ completed    English     <NA>    06/14/~ 0.70240~ 06/14/2019 0.740~
## 6 SNAP 2015~ completed    English     <NA>    06/14/~ 0.73562~ 06/14/2019 0.745~
## # ... with 266 more variables: ID.time <chr>, Q1 <chr>, Q1.a <chr>, Q2 <chr>,
## #   Q3 <chr>, Q3.a <chr>, Q4 <chr>, Q5 <chr>, Q6 <chr>, Q6.a <chr>, Q7 <chr>,
## #   Q7.a <chr>, Q8 <chr>, Q8.a <chr>, Q9 <chr>, Q9.a <chr>, Q10 <chr>,
## #   Q10.a <chr>, Q11 <chr>, Q11.a <chr>, Q12:1 <chr>, Q12:2 <chr>, Q12:3 <chr>,
## #   Q12:4 <chr>, Q12:5 <chr>, Q12:6 <chr>, Q12:7 <chr>, Q12:8 <chr>,
## #   Q12:9 <chr>, Q12:10 <chr>, Q12:11 <chr>, Q12.a <chr>, Q13 <chr>, Q14 <chr>,
## #   Q14.a <chr>, Q15.a <chr>, Q15.b <chr>, Q15.c <chr>, Q15.d <chr>, ...
```

DATA WRANGLIN' AN' CLEANIN' Creation of ancillary duplicate data frames

```
df1 <- dfog
df2 <- df1
df3 <- df2
```

Remove the top row and use the second row (the questions themselves) as the column identifier. The
semicolons and periods in the top row creates many issues

```
df4<-df3
names(df4) <- as.matrix(df4[1, ])
```

```
## Warning: The 'value' argument of 'names<-' can't be empty as of tibble 3.0.0.
## Column 123 must be named.
```

```
df4 <- df4[-1, ]
df4[] <- lapply(df4, function(x) type.convert(as.character(x)))
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

```
## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE

## Warning in type.convert.default(as.character(x)): 'as.is' should be specified by
## the caller; using TRUE
```

remove all columns where all values are NA

```
df5 <- df4
df6a <- df5[ , colSums(is.na(df4)) < nrow(df4)]
```

Rename columns that currently have the same name, as that causes issues. Rename column 117 name in df6a, because the column name in that position is blank. I am not sure how the bottom portion works with attributing df6 to df6a after I declared df6, but it does. I tried it other ways and it didn't output df6. To test this, please do comment out df6 <- df6a and test yourself. I use this work flow to rename all repeated bits, because doing it any other way wont work. dplyr's tools didnt work, renaming was only accomplished with base r.

```
df6 <- df6a
df6 <- names(df6a)[117] <- "How much do you expect your PhD to improve your job prospects?"     ###AV C


###AV Rename all duplicates of "If other, please specify" and make unique
df6 <- names(df6a)[12] <- "specified1"
df6 <- names(df6a)[16] <- "specified2"
df6 <- names(df6a)[22] <- "specified3"

df6 <- names(df6a)[34] <- "specified4"
df6 <- names(df6a)[37] <- "specified5"
df6 <- names(df6a)[54] <- "specified6"
df6 <- names(df6a)[79] <- "specified7"
df6 <- names(df6a)[89] <- "specified8"

df6 <- names(df6a)[104] <- "specified9"
df6 <- names(df6a)[116] <- "specified10"
df6 <- names(df6a)[139] <- "specified11"
df6 <- names(df6a)[141] <- "specified12"
df6 <- names(df6a)[146] <- "specified13"

df6 <- names(df6a)[158] <- "specified14"
df6 <- names(df6a)[173] <- "specified15"
df6 <- names(df6a)[183] <- "specified16"
df6 <- names(df6a)[191] <- "specified17"
df6 <- names(df6a)[200] <- "specified18"

df6 <- names(df6a)[225] <- "specified19"
df6 <- names(df6a)[233] <- "specified20"
df6 <- names(df6a)[239] <- "specified21"
df6 <- names(df6a)[255] <- "specified22"
df6 <- names(df6a)[261] <- "specified23"
```

```r
###AV Rename all duplicates of "What prompted you to study outside your country of upbringing" and make
df6 <- names(df6a)[23] <- "prompt to study outside country of upbringing1"
df6 <- names(df6a)[24] <- "prompt to study outside country of upbringing2"
df6 <- names(df6a)[25] <- "prompt to study outside country of upbringing3"
df6 <- names(df6a)[26] <- "prompt to study outside country of upbringing4"
df6 <- names(df6a)[27] <- "prompt to study outside country of upbringing5"
df6 <- names(df6a)[28] <- "prompt to study outside country of upbringing6"
df6 <- names(df6a)[29] <- "prompt to study outside country of upbringing7"
df6 <- names(df6a)[30] <- "prompt to study outside country of upbringing8"
df6 <- names(df6a)[31] <- "prompt to study outside country of upbringing9"
df6 <- names(df6a)[32] <- "prompt to study outside country of upbringing10"
df6 <- names(df6a)[33] <- "prompt to study outside country of upbringing11"


###AV Rename all duplicates of "Who was the perpetrator(s)" and make unique
df6 <- names(df6a)[97] <-  "positionOfPerpetrator1"
df6 <- names(df6a)[98] <-  "positionOfPerpetrator2"
df6 <- names(df6a)[99] <-  "positionOfPerpetrator3"
df6 <- names(df6a)[100] <- "positionOfPerpetrator4"
df6 <- names(df6a)[101] <- "positionOfPerpetrator5"
df6 <- names(df6a)[102] <- "positionOfPerpetrator6"
df6 <- names(df6a)[103] <- "positionOfPerpetrator7"


###AV Rename all duplicates of "Which of the following have you experienced?" and make unique
df6 <- names(df6a)[107] <- "experience1"
df6 <- names(df6a)[108] <- "experience2"
df6 <- names(df6a)[109] <- "experience3"
df6 <- names(df6a)[110] <- "experience4"
df6 <- names(df6a)[111] <- "experience5"
df6 <- names(df6a)[112] <- "experience6"
df6 <- names(df6a)[113] <- "experience7"
df6 <- names(df6a)[114] <- "experience8"
df6 <- names(df6a)[115] <- "experience9"


###AV Rename all duplicates of "If you're unlikely to pursue an academic research career, what are the r
df6 <- names(df6a)[130] <- "reasonUnlikelyAcademicCarreerPursuit1"
df6 <- names(df6a)[131] <- "reasonUnlikelyAcademicCarreerPursuit2"
df6 <- names(df6a)[132] <- "reasonUnlikelyAcademicCarreerPursuit3"
df6 <- names(df6a)[133] <- "reasonUnlikelyAcademicCarreerPursuit4"
df6 <- names(df6a)[134] <- "reasonUnlikelyAcademicCarreerPursuit5"
df6 <- names(df6a)[135] <- "reasonUnlikelyAcademicCarreerPursuit6"
df6 <- names(df6a)[136] <- "reasonUnlikelyAcademicCarreerPursuit7"
df6 <- names(df6a)[137] <- "reasonUnlikelyAcademicCarreerPursuit8"
df6 <- names(df6a)[138] <- "reasonUnlikelyAcademicCarreerPursuit9"

###AV Rename all duplicates of "How did you arrive at your current career decision? Plea..." and make u
df6 <- names(df6a)[147] <- "HowArriveAtCareerDecision1"
df6 <- names(df6a)[148] <- "HowArriveAtCareerDecision2"
df6 <- names(df6a)[149] <- "HowArriveAtCareerDecision3"
df6 <- names(df6a)[151] <- "HowArriveAtCareerDecision4"
df6 <- names(df6a)[152] <- "HowArriveAtCareerDecision5"
```

```r
df6 <- names(df6a)[153] <- "HowArriveAtCareerDecision6"
df6 <- names(df6a)[154] <- "HowArriveAtCareerDecision7"
df6 <- names(df6a)[155] <- "HowArriveAtCareerDecision8"
df6 <- names(df6a)[156] <- "HowArriveAtCareerDecision9"
df6 <- names(df6a)[157] <- "HowArriveAtCareerDecision10"

###AV Rename all duplicates of "How do you learn about available career opportunities that are beyond a
df6 <- names(df6a)[159] <- "HowLearnCareerNotAcademia1"
df6 <- names(df6a)[160] <- "HowLearnCareerNotAcademia2"
df6 <- names(df6a)[161] <- "HowLearnCareerNotAcademia3"
df6 <- names(df6a)[162] <- "HowLearnCareerNotAcademia4"
df6 <- names(df6a)[163] <- "HowLearnCareerNotAcademia5"
df6 <- names(df6a)[164] <- "HowLearnCareerNotAcademia6"
df6 <- names(df6a)[165] <- "HowLearnCareerNotAcademia7"
df6 <- names(df6a)[166] <- "HowLearnCareerNotAcademia8"
df6 <- names(df6a)[167] <- "HowLearnCareerNotAcademia9"
df6 <- names(df6a)[168] <- "HowLearnCareerNotAcademia10"
df6 <- names(df6a)[169] <- "HowLearnCareerNotAcademia11"
df6 <- names(df6a)[170] <- "HowLearnCareerNotAcademia12"
df6 <- names(df6a)[171] <- "HowLearnCareerNotAcademia13"
df6 <- names(df6a)[172] <- "HowLearnCareerNotAcademia14"

###AV Rename all duplicates of "Which of the following 3 things would you say are the most difficult for
df6 <- names(df6a)[174] <- "DifficultInDiscipline1"
df6 <- names(df6a)[175] <- "DifficultInDiscipline2"
df6 <- names(df6a)[176] <- "DifficultInDiscipline3"
df6 <- names(df6a)[177] <- "DifficultInDiscipline4"
df6 <- names(df6a)[178] <- "DifficultInDiscipline5"
df6 <- names(df6a)[179] <- "DifficultInDiscipline6"
df6 <- names(df6a)[180] <- "DifficultInDiscipline7"
df6 <- names(df6a)[181] <- "DifficultInDiscipline8"
df6 <- names(df6a)[182] <- "DifficultInDiscipline9"


###AV Rename all duplicates of "Which of the following would you say are the most difficult for PhD stu
df6 <- names(df6a)[184] <- "DifficultInCountry1"
df6 <- names(df6a)[185] <- "DifficultInCountry2"
df6 <- names(df6a)[186] <- "DifficultInCountry3"
df6 <- names(df6a)[187] <- "DifficultInCountry4"
df6 <- names(df6a)[188] <- "DifficultInCountry5"
df6 <- names(df6a)[189] <- "DifficultInCountry6"
df6 <- names(df6a)[190] <- "DifficultInCountry7"

###AV Rename all duplicates of "Which of the following resources do you think PhD students need the most
df6 <- names(df6a)[192] <- "ResourcesForSatisfyingCareer1"
df6 <- names(df6a)[193] <- "ResourcesForSatisfyingCareer2"
df6 <- names(df6a)[194] <- "ResourcesForSatisfyingCareer3"
df6 <- names(df6a)[195] <- "ResourcesForSatisfyingCareer4"
df6 <- names(df6a)[196] <- "ResourcesForSatisfyingCareer5"

df6 <- names(df6a)[197] <- "ResourcesForSatisfyingCareer6"
df6 <- names(df6a)[198] <- "ResourcesForSatisfyingCareer7"
df6 <- names(df6a)[199] <- "ResourcesForSatisfyingCareer8"
```

```r
###AV Rename all duplicates of "Which, if any, of the following activities have you done to advance you
df6 <- names(df6a)[217] <- "ActivitiesToAdvanceCareer1"
df6 <- names(df6a)[218] <- "ActivitiesToAdvanceCareer2"
df6 <- names(df6a)[219] <- "ActivitiesToAdvanceCareer3"
df6 <- names(df6a)[220] <- "ActivitiesToAdvanceCareer4"
df6 <- names(df6a)[221] <- "ActivitiesToAdvanceCareer5"

df6 <- names(df6a)[222] <- "ActivitiesToAdvanceCareer6"
df6 <- names(df6a)[223] <- "ActivitiesToAdvanceCareer7"
df6 <- names(df6a)[224] <- "ActivitiesToAdvanceCareer8"


###AV Rename all duplicates of "Which of the following social media networks have you used to build you
df6 <- names(df6a)[226] <- "SocialMediaToBuildNetwork1"
df6 <- names(df6a)[227] <- "SocialMediaToBuildNetwork2"
df6 <- names(df6a)[228] <- "SocialMediaToBuildNetwork3"
df6 <- names(df6a)[229] <- "SocialMediaToBuildNetwork4"
df6 <- names(df6a)[230] <- "SocialMediaToBuildNetwork5"
df6 <- names(df6a)[231] <- "SocialMediaToBuildNetwork6"
df6 <- names(df6a)[232] <- "SocialMediaToBuildNetwork7"

###AV Rename all duplicates of "What would you do differently right now if you were starting your progra
df6 <- names(df6a)[234] <- "DoDifferently1"
df6 <- names(df6a)[235] <- "DoDifferently2"
df6 <- names(df6a)[236] <- "DoDifferently3"
df6 <- names(df6a)[237] <- "DoDifferently4"
df6 <- names(df6a)[238] <- "DoDifferently5"


###AV Rename all duplicates of "Which of the following best describes you?" and make unique
df6 <- names(df6a)[243] <- "BestDescribes1"
df6 <- names(df6a)[244] <- "BestDescribes2"
df6 <- names(df6a)[245] <- "BestDescribes3"
df6 <- names(df6a)[246] <- "BestDescribes4"
df6 <- names(df6a)[247] <- "BestDescribes5"

df6 <- names(df6a)[248] <- "BestDescribes6"
df6 <- names(df6a)[249] <- "BestDescribes7"
df6 <- names(df6a)[250] <- "BestDescribes8"
df6 <- names(df6a)[251] <- "BestDescribes9"
df6 <- names(df6a)[252] <- "BestDescribes10"

df6 <- names(df6a)[253] <- "BestDescribes11"
df6 <- names(df6a)[254] <- "BestDescribes12"


###AV Rename all duplicates of "Do you have any caring responsibilities" and make unique
df6 <- names(df6a)[256] <- "CaringResponsibilities1"
df6 <- names(df6a)[257] <- "CaringResponsibilities2"
df6 <- names(df6a)[258] <- "CaringResponsibilities3"
df6 <- names(df6a)[259] <- "CaringResponsibilities4"
df6 <- names(df6a)[260] <- "CaringResponsibilities5"
```

```
###AV Final Important Bits, one more rename and the declaration that makes renaming work.
df6 <- names(df6a)[242] <- "Gender"
df6 <- df6a

###AV Workflow Notes
###AV This bit told me what I am missing to rename and make not identical
###rename(df6, combine1 = 150)

###AV This is the template I used to fill in the above.
####AV Rename all duplicates of "" and make unique
#df6 <- names(df6a)[] <- ""
#df6 <- names(df6a)[] <- ""
#df6 <- names(df6a)[] <- ""
#df6 <- names(df6a)[] <- ""
#df6 <- names(df6a)[] <- ""



###Finished Product: df7
df7 <- df6
```

Combine demographic data

```
df8 <- unite(df7, "ethnicityGroupedOther",  243:254, remove = FALSE, na.rm = TRUE)
#df7 %>% unite("ethnicityOtherWriteInIncluded", 2,remove= FALSE) specified22 is variable that has the w
df9 <- unite(df8, "ethnicityOtherWriteInNotYetIncluded", 244:252, remove = FALSE, na.rm = TRUE)
###AV Ensure that all NA values remain as NA, as the above unite processes seem to convert NA to blanks
df10 <- df9
df10[df10 == ""] <- NA
###AV finally combine them all
df11 <- unite(df10, "ethnicityOtherWriteInIncluded", c('ethnicityOtherWriteInNotYetIncluded', 'specified
```

—Everything Below this line probably requires df12 or some variation of df11—

Well, doing the above combination wasnt quite that useful, because now I will focus on PEOPLE WHO MOVED INTO A COUNTRY FROM ELSEWHERE AND GENDER. Not on ethnicity. But this is EDA, so the clear and distinct lack of direction is a feature, not a bug.

Gender Analysis against Training and Readiness (Q50)

Gender count and gender proportion analysis, and data frame creation

```
###Output A Table and dataframe with the values needed
tableGender <- table(df11['Gender'])
genderCount<- as.data.frame(tableGender)

###Proportional Analysis
genderProportion <- as.data.frame(table(df11$Gender)/length(df11$Gender))
###Display Output genderProportion
genderProportion
```

```
##                               Var1        Freq
## 1 Female (including trans female) 0.499706400
## 2         Gender queer / Non binary 0.004697592
## 3      Male (including trans male) 0.489136817
## 4                  Prefer not to say 0.006459190
```

Creation of different data frames for analysis by gender, against training efficacy for certain tasks(Q50).

```
df12 <- df11[c(201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,242)]

df12M <- df12[df12$'Gender' == 'Male (including trans male)',]
df12F <- df12[df12$'Gender' == 'Female (including trans female)',]
df12GQNB <- df12[df12$'Gender' == 'Gender queer / Non binary',]
```

Whole workflow for analysis using "Male (including trans male)" value row selection

```
###Select relevant columns/variables

### Rename Variable/Column to be analyzed, the dependent, by index
df13M <- df12M
df14M <- df13M
df14M <- names(df13M)[1] <- "Collecting"
df14M <- df13M


###Plot the bit
#ggplot(df14M, aes( Collecting, colour = Gender)) +   geom_bar() ###Just another way to plot the data

barplotMCollectingData <- ggplot(df14M, aes( Collecting, colour = Gender)) +
  geom_bar()+
  scale_x_discrete(guide = guide_axis(n.dodge=2))

###Display Plot
barplotMCollectingData
```
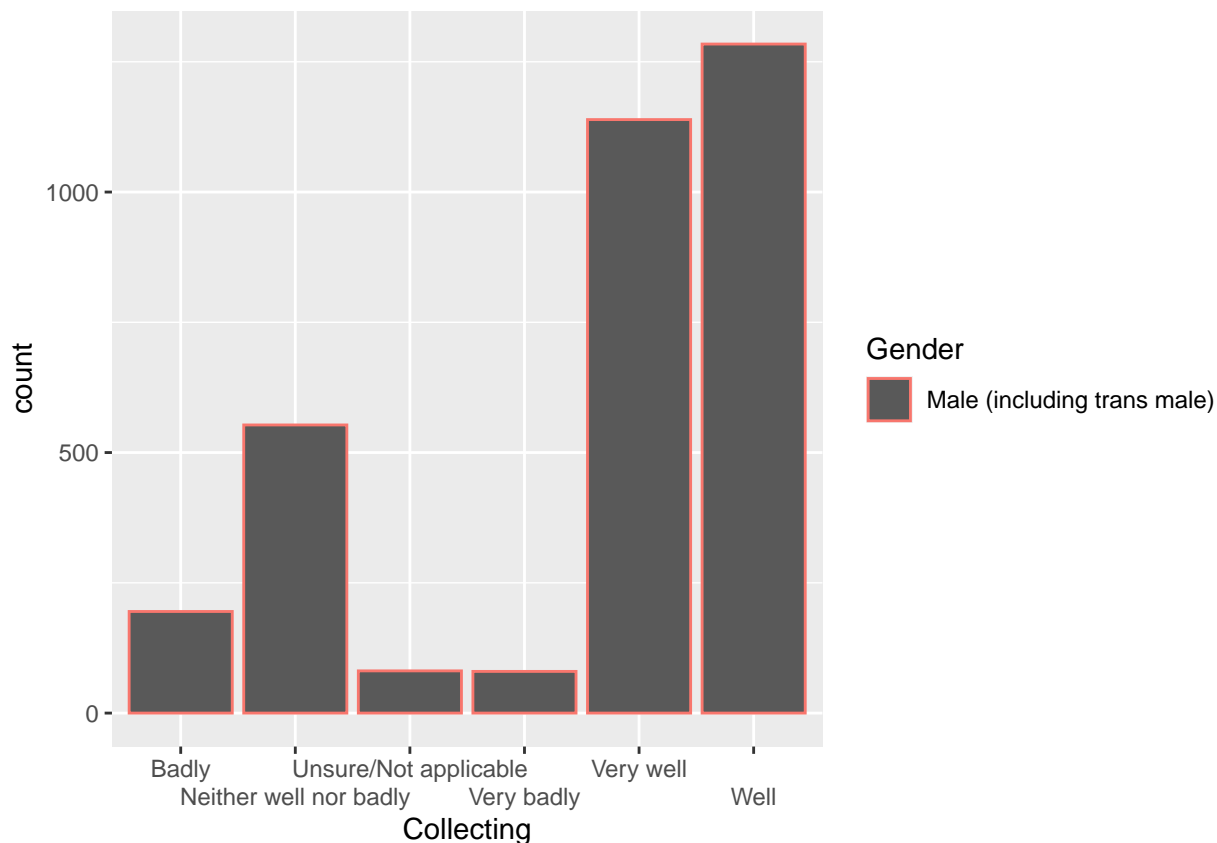
Analysis for Collecting Data Variable against Gender, Count and Proportion

```
###Output A Table and dataframe, possible first step in analysis given grouping. difficult to say, would
#Collecting <- table(df12['Collecting'])
#genderVsCollecting <- as.data.frame.array(Collecting)
#genderVsCollecting


###Proportional Analysis
CollectingProportion <- as.data.frame(table(df12$Collecting)/length(df12$Collecting))
```

```
## Warning: Unknown or uninitialised column: 'Collecting'.
```

```
## Warning: Unknown or uninitialised column: 'Collecting'.
```

```
###Display Output
CollectingProportion
```

```
## [1] table(df12$Collecting)/length(df12$Collecting)
## <0 rows> (or 0-length row.names)
```

Some data transformations and a rename, so that I could copy and paste from online sources which predominantly use "df" or "data" and not df23 or whatever I'm doing. Here is where I seem to have renamed df12's Collecting column and not df12m's, accounting for the difference between the two.

```
df15 <- names(df12)[1] <- "Collecting"
df15 <- df12
df <- df15
```

Graphs!!!

####Gender against Collecting Data

first graph, not incredibly interesting and also hampered by small sample size of Gender Queer/Non Binary and Prefer Not To Say

```
groupedBarGenderCollecting <- ggplot(data = df) +
  geom_bar(mapping = aes(x = Gender, fill = Collecting), position = "dodge")+
  scale_colour_brewer(palette = "BrBG") +
  scale_fill_brewer(palette = "BrBG") +
  scale_x_discrete(guide = guide_axis(n.dodge=2))


groupedBarGenderCollecting
```
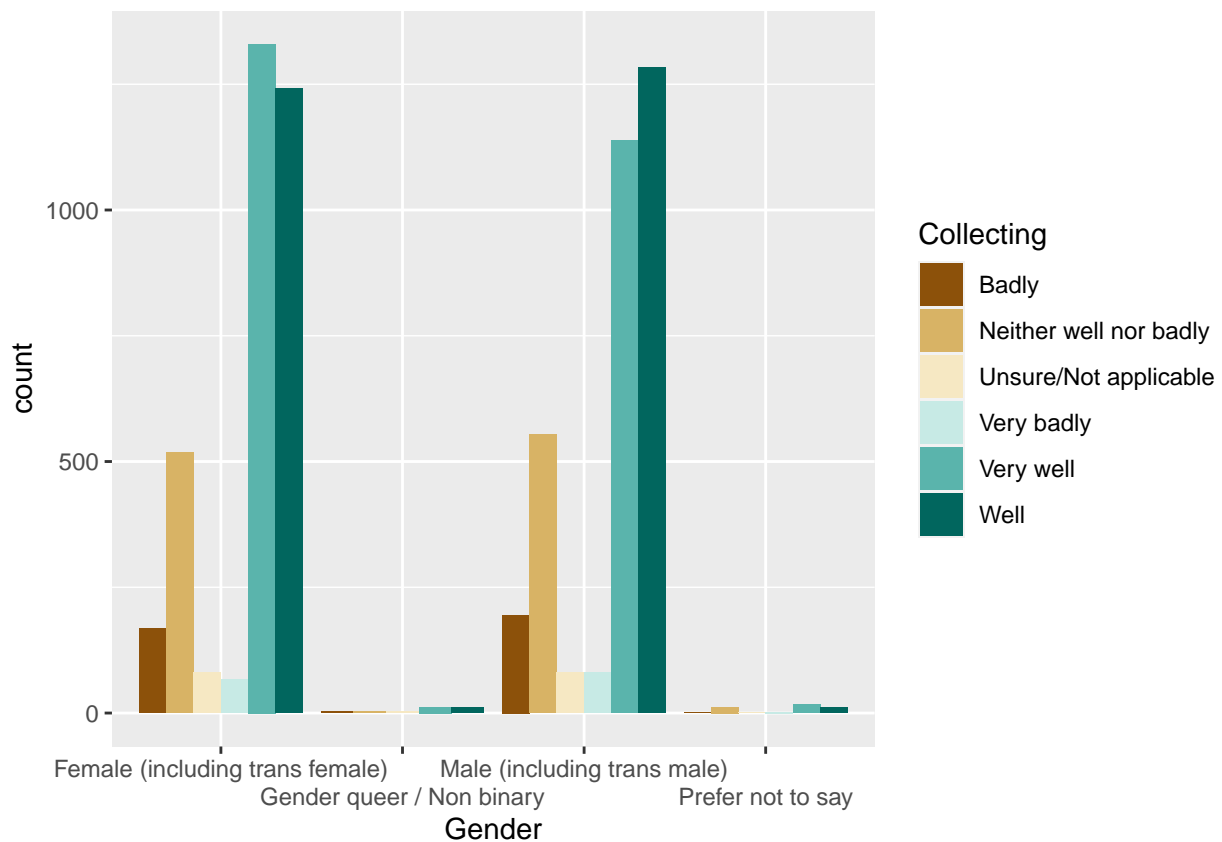

A better proportional graph. This chunk also contains the code to reorder variables manually.

```
data_new <- df
df <- data_new
```

```
df$Collecting <- factor(df$Collecting , levels=c("Very badly", "Badly", "Neither well nor badly", "Well"

data_new <- df
df <- data_new


ggplot(data = data_new) +
  geom_bar(mapping = aes(x = Gender, fill = Collecting), position = "fill")+
  scale_colour_brewer(palette = "BrBg") +
  scale_fill_brewer(palette = "BrBG") +
  scale_x_discrete(guide = guide_axis(n.dodge=2))
```

Warning in pal_name(palette, type): Unknown palette BrBg
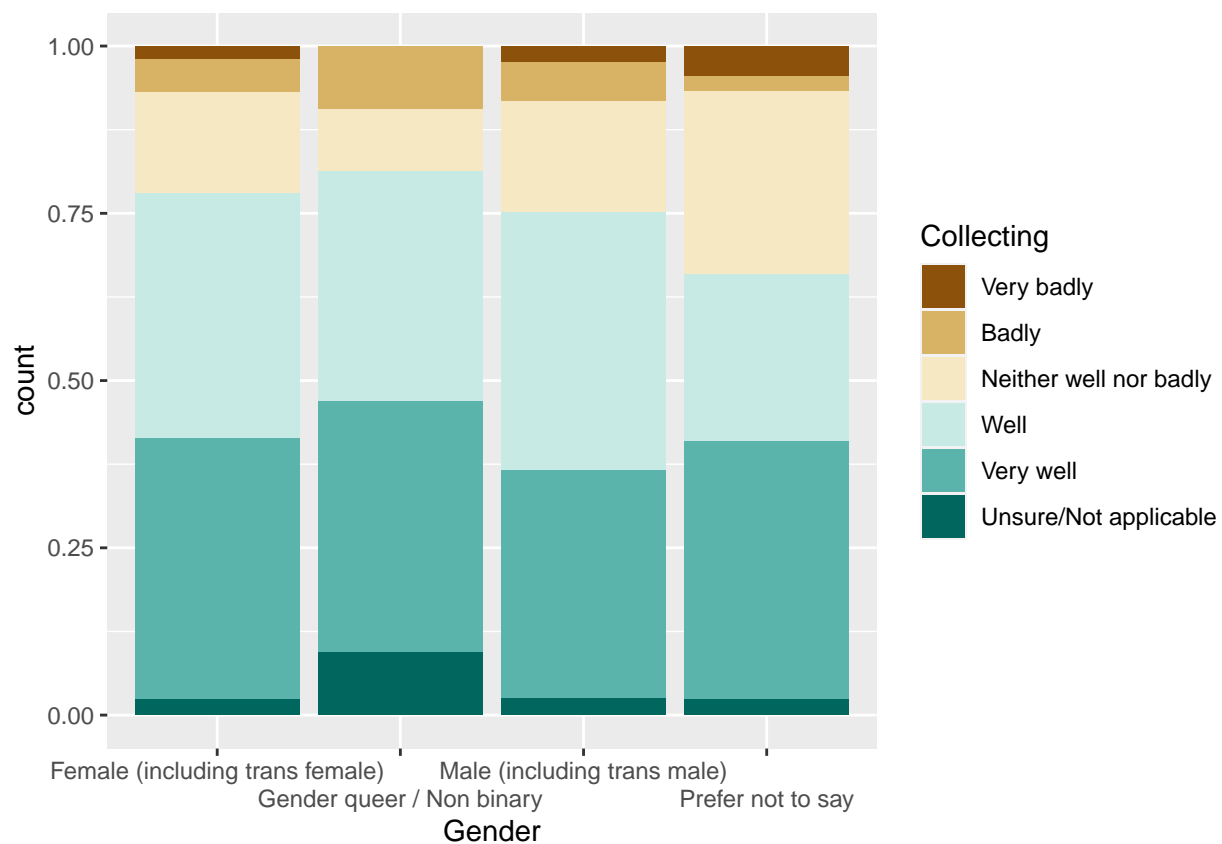
Not bad, however the not sure/not applicable bit is distorting our visual analysis.

Same proportional graph, without Not Applicable variable value

```
###New data frame, to ensure there are no issues down the line
graphdf1gender <- df
###Keep all values that are not containing "Unsure/Not applicable"
data_new1 <- graphdf1gender[!(graphdf1gender$Collecting == "Unsure/Not applicable"),]

###Refresh old graphdf1gender with data_new1 info
graphdf1gender <- data_new1
```

```
###Set variable order
graphdf1gender$Collecting <- factor(graphdf1gender$Collecting , levels=c("Very badly", "Badly", "Neither

##Plot
ggplot(data = data_new1) +
  geom_bar(mapping = aes(x = Gender, fill = Collecting), position = "fill")+
  scale_colour_brewer(palette = "BrBg") +
  scale_fill_brewer(palette = "BrBG") +
  scale_x_discrete(guide = guide_axis(n.dodge=2))
```

Warning in pal_name(palette, type): Unknown palette BrBg
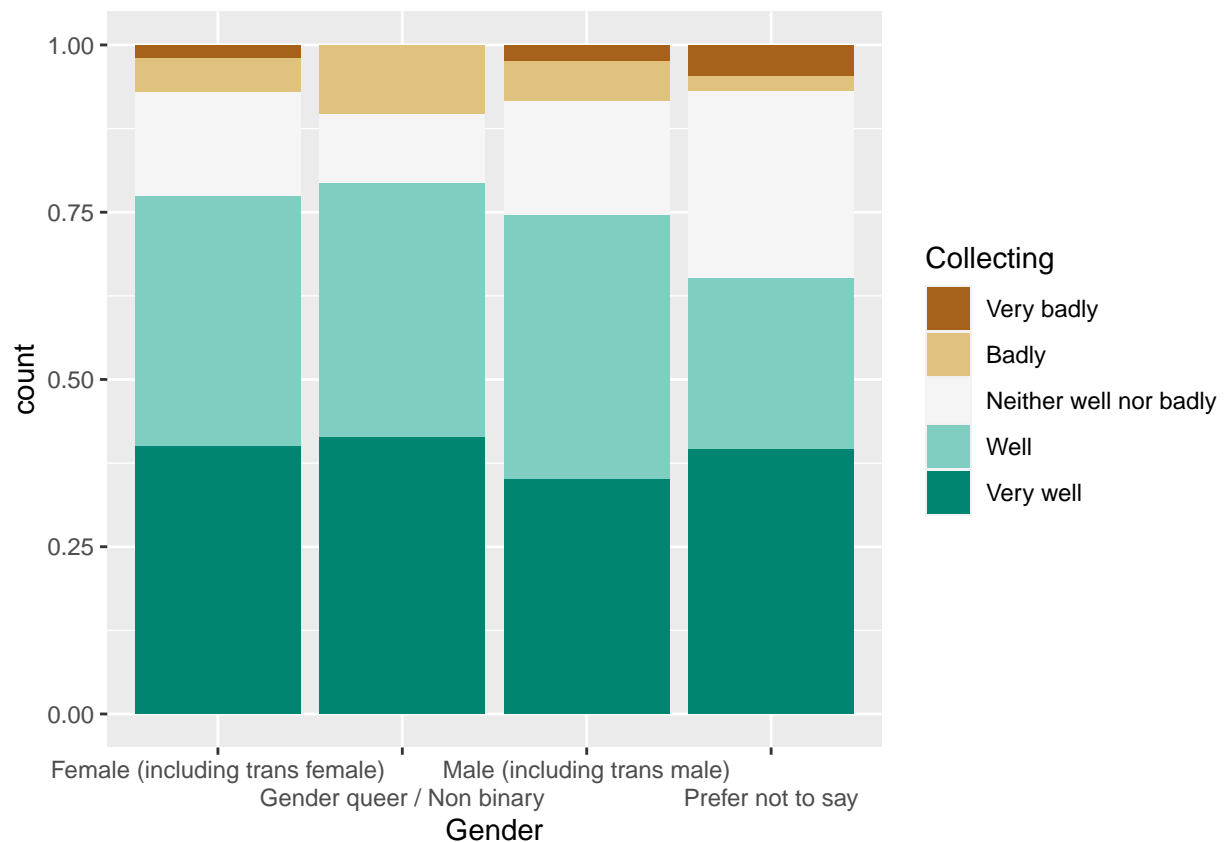
Gender Against Analyzing Data

Gender Against Designing robust reproducible experiments

Gender Against Writing a paper for publication in a peer-reviewed journal

Gender Against Developing resilience to manage rejection by a peer review panel

Gender Against Presenting findings to a specialist audience

Gender Against Presenting findings to a non-specialist (public) audience

Gender Against Finding a satisfying career

Gender Against Managing complex projects

Gender Against Developing a business plan

Gender Against Managing people

Gender Against Managing a large operational budget

Gender Against I feel that my programme is preparing me well for a research career

Gender Against I feel that my programme is perparing me well for a non-research science-related career

Gender Against I feel that my programme is preparing me well for a career that straddles both industry and academia   This will be the bread and butter of this assignment. It looks beautiful. Waffle Graph, must have info input manually. Still beautiful.

```
parts <- c(`Un-breached\nUS Population` = (318 - 11 - 79), `Premera` = 11, `Anthem` = 79)

waffle(
  parts, rows = 8, size = 1,
  colors = c("#969696", "#1879bf", "#009bda"), legend_pos = "bottom"
)
```
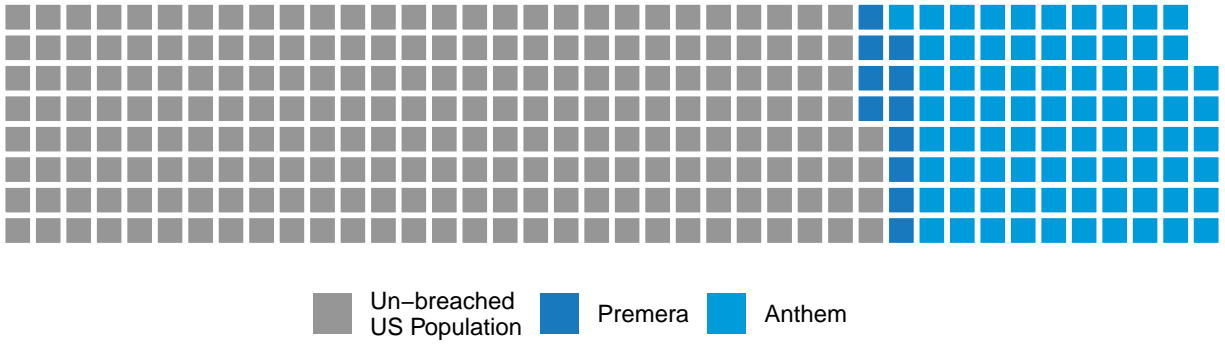
Analysis that isn't working yet

```
#lm1 <- lm(data = df, Collecting ~ Gender) # the model

#summary(lm1) # summarizes the output of the model
```

Appendix: Misc Example Code Area

How to rename once

```
###AV Script to rename one or multiples, same for multiples but just repeat first line till the end and

#df14 <- names(df13)[1] <- "Collecting"
#df14 <- df13
```

Basic plot

```
#ggplot(data = df) +
    ###actual plot bit
 # geom_bar(mapping = aes(x = Gender, fill = Collecting), position = "fill")+
    ###color
  #scale_colour_brewer(palette = "BrBg") +
   #scale_fill_brewer(palette = "BrBG") +
  ### Makes sure that the names/titles/labels do not overlap
     #scale_x_discrete(guide = guide_axis(n.dodge=2))
```

Creating new data frames by selecting only rows with a certain value

```
#Possibly create new dataframes by row value
#df12Male <- df12[df12$'Gender' == 'Male (including trans male)',]
#df12Female <- df12[df12$'Gender' == 'Female (including trans female)',]
#df12GenderQueerandorNonBinary <- df12[df12$'Gender' == 'Gender queer / Non binary',]
```