**Dublin Business School**

**Electronic Assignment Cover sheet**

**Student (s) Number as per your student card:** Juliana Salvadori - No: 10521647

**Course Title:** MSc Data Analytics

**Lecturer Name:** Terri Hoare

**Module/Subject Title:** Data Visualisation (B9DA106)

**Assignment Title:** CA2 - Summative assessment - EDA

GitHub Python code:

<https://gist.github.com/jusalvadori/67cdb4b6d1f3867ce519b4c323e4c46b>

GitHub R code:

<https://gist.github.com/jusalvadori/ba15b368b0e1eeb3309b34d88494c8f8>

**Data Visualisation - Exploratory Data Analysis**

In this task we have explored a dataset based on a survey called ‘*What Do Men Think It Means To Be A Man?*’ from FiveThirtyEight website [1], to apply some of the visualisation techniques learned on this module. This study was carried out at a time when lots of questions arose regarding claims of sexual harassment, gender differences in the workplace and the [role of masculinity in society](https://www.nbcnews.com/think/opinion/father-s-day-men-are-experiencing-crisis-masculinity-solution-more-ncna884051), and it aimed to understand how men fell about being a man and what they think about masculinity. The dataset has 1,615 adults’ answers, who identified themselves as men, to nearly 40 questions about masculinity, workplace culture and intimacy, and other related subjects. Hence, we selected a few questions from the survey to perform an exploratory data analysis of the data using three different visualisation tools: Tableau, R and Python.

Tableau is a visualisation tool utilized to help on data exploration and analysis that does not require any type of technical or programming skills as it presents the data in a simplified format that is easy to understand and work with [10]. Tableau is broadly used for data analytics and allows you to build from a simple chart, reports to complex dashboards. It works similarly to Excel at certain points and that is one of the reasons it is easy to operate, as many users already have some familiarity with Excel, however, Tableau is more powerful for visualisation purpose.

R is a programming language specific for statists purpose and it has some basic visualisation package by default with a few functions that allows you to create the most common plots [8]. However, there are other more advanced packages such as ggplot2, lattice, plotly and shiny, that allows you to work with different types of plots and provide a range of functions to improve plots appearance and interactivity. For this task we have used the ggplot2 library.

Python is also a programming language; however, it is a general-purpose one, and widely used for data analysis due to its capacity to deal with large datasets [6]. As Python is not a specific statistic or visualisation tool it does not have any visualisation package by default but has several visualisation libraries available such as matplotlib, ggplot and seaborn, that can be easily imported and used. For this task we have used the matplotlib and seaborn libraries.

Both R and Python, unlike Tableau, require some technical and/or programming skills to be operated and can take longer to build a simple chart as you might need to apply some data manipulation to get it in the right format to be able to call a plot function. For instance, Python and R do not have an easy way to add values or percentages to each part of the chart, you need to add a few lines to the code to do that and the percentages need to be calculated beforehand and then added to the chart. In Tableau, there is a more straightforward way to do that with a few clicks.

The dataset is mainly made up of categorical data, as most of the questions have a limited set of possible answers and therefore we have selected a few types of bar and line chart to show the distribution of data by age group, civil status, and sexual orientation.

Firstly, we had a look at the overall distribution of the respondents, as shown in Figure 1, in Tableau is easier and clearer to show that information using simple reporting tables and horizontal bar chart.

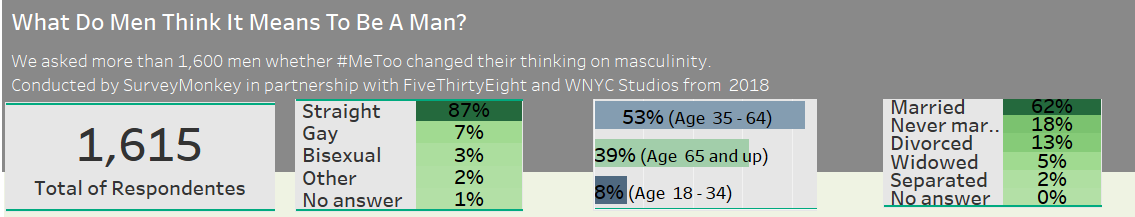


Figure 1 - Tableau

In Figure 2 and Figure 4 we show the same distribution in Python and Figure 3 and Figure 5 in R using a few different types of charts. Despite of the recommendations to not use or avoid using pie charts, it can be cautiously used to show a part-to-whole comparison when there are only a few categories [4] and therefore we used that type to build the first chart showing the total of respondents.

For Total respondents by age group, both in R and in Python, we have used horizontal bar chart, as it is one of the easiest methods to visualize categorical data comparisons [5]. Note that, as there are only 3 groups, we can quickly understand the distribution, even without using a range of different colours as shown in Figure 3.

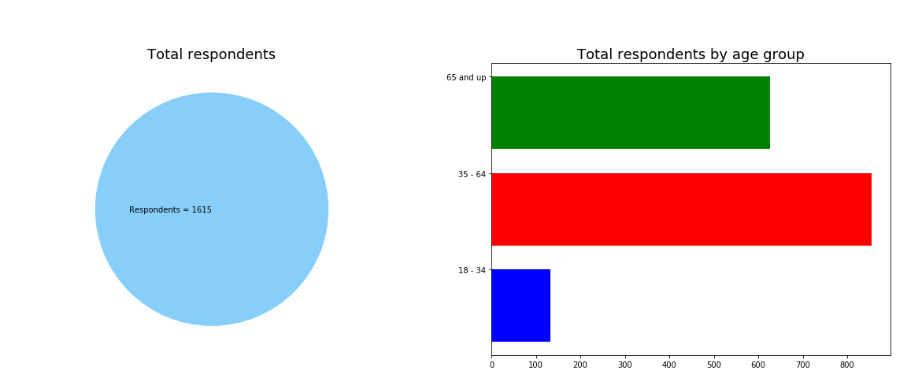


Figure 2 – Python

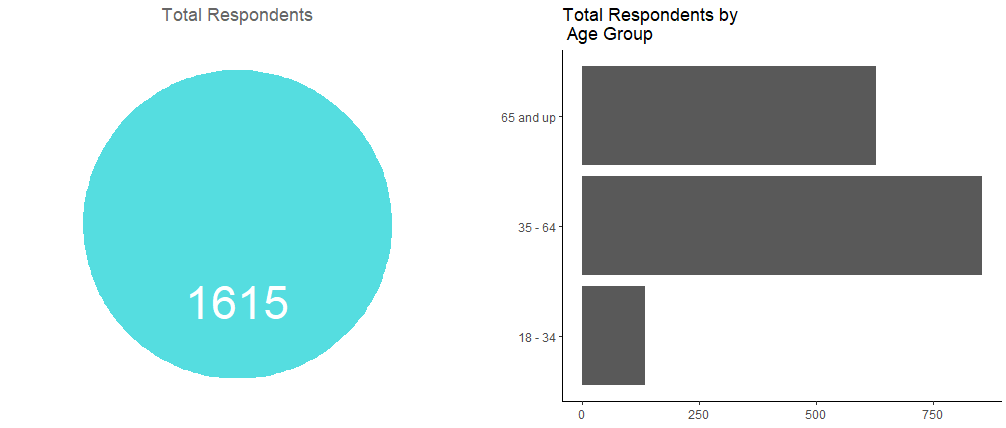


Figure 3 – R

The line chart is frequently used to visualize trends over time [5], however, we decided to use that chart to show the total number of respondents by civil status to see how that type of chart looks like on comparisons, as there is no time-series data on the dataset. It worked fine for this purpose, we can clearly identify that most of the respondents is married followed by divorced and never married men. Also, it is worth noting how the presentation of that data in Figure 5 looks much better than in Figure 4 when using the background grid in a light shade of grey that does not blur the data representation and the thicker line to join the data points.

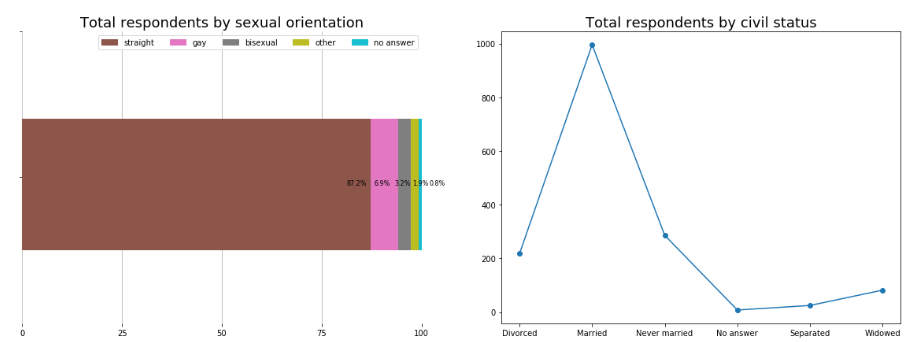


Figure 4 - Python

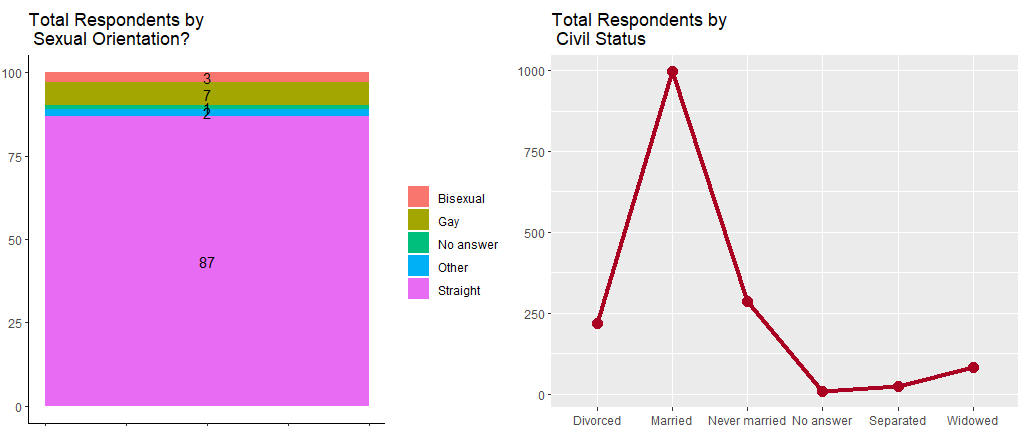


Figure 5 - R

Next, we explored some of the questions answered on the survey, starting for “*In general, how masculine or “manly” do you feel?*”. In Tableau, we represented those answers using a line chart shading the area under the lines to facilitate the understanding of the contribution of that line to the whole, as it is a better way to use that type of chart for comparisons between categories.

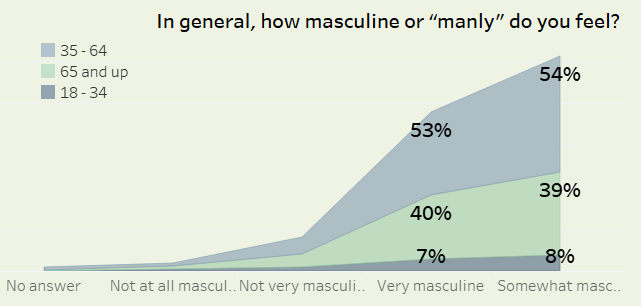


Figure 6 - Tableau

In Python and in R, we decided to go with bar charts as shown in Figure 7 and Figure 8. For both cases, the choice of colours worked fine as there are only three different age groups, and we can easily identify that most of the answers for each group was that the respondent feels somewhat masculine. Observe that in Figure 9 and Figure 10 the same answers distributed by civil status start to become difficult to understand as there are more categories to analyse and, therefore, many colours to visualise.

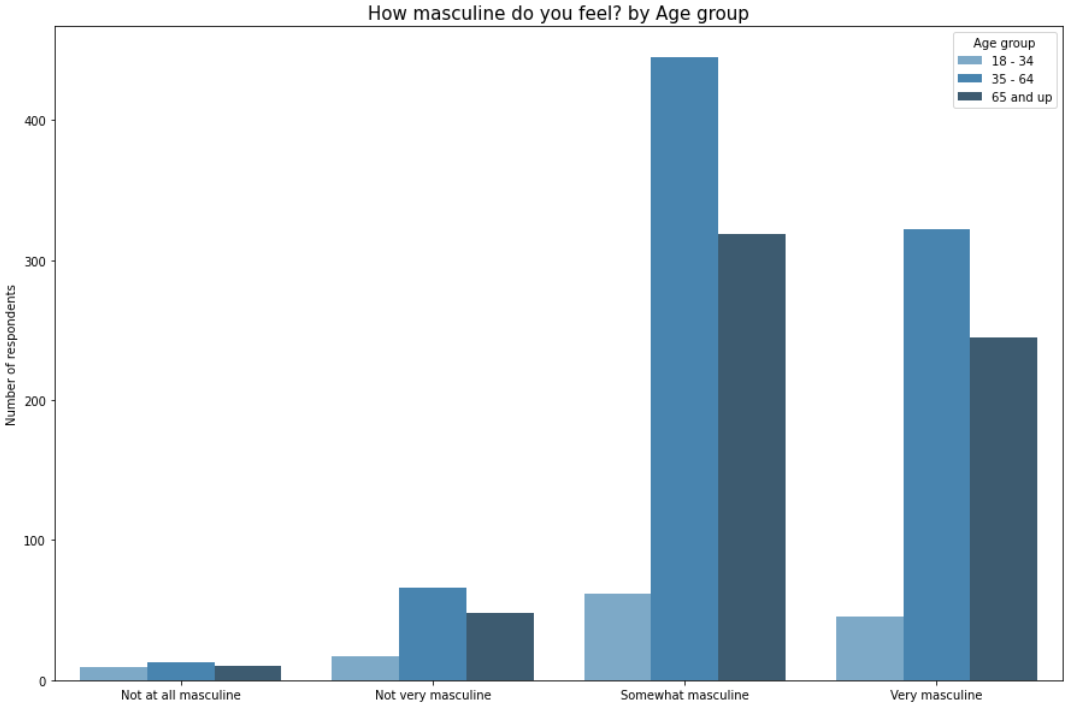


Figure 7 – Python

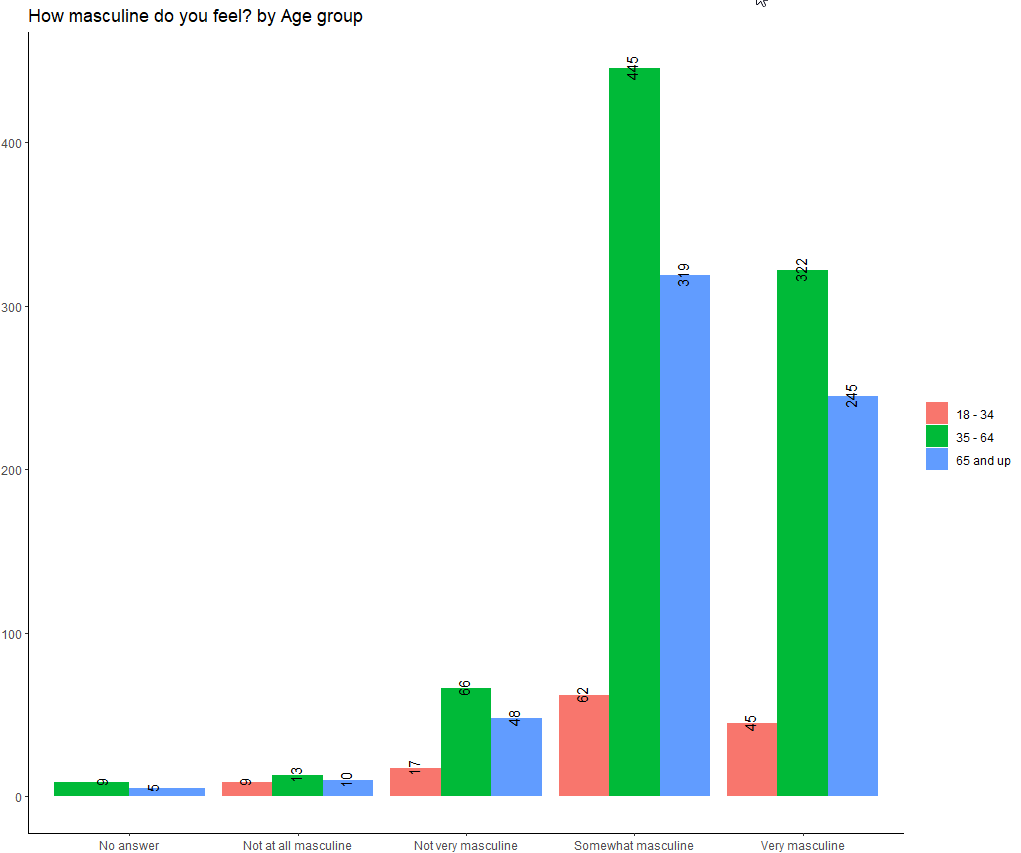


Figure 8 - R

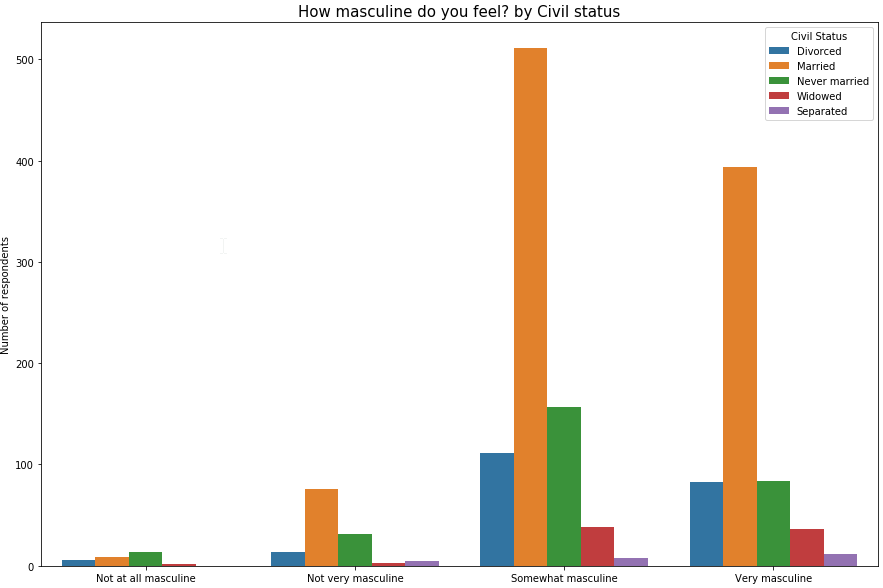


Figure 9 – Python

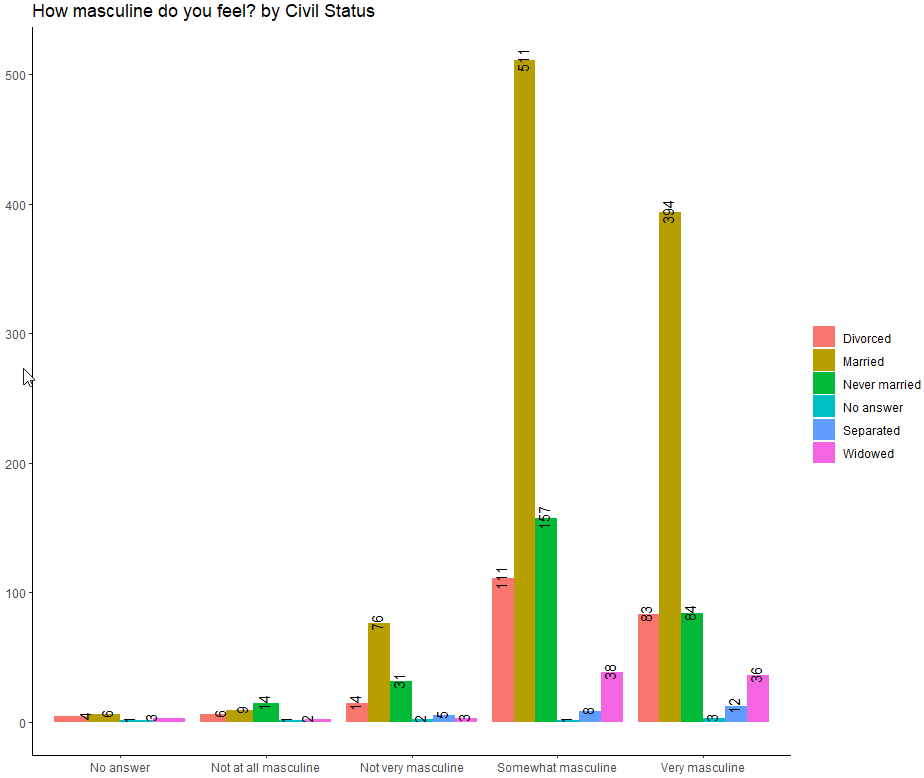


Figure 10 - R

For the next question analysed “*Do you think that society puts pressure on men in a way that is unhealthy or bad for them?*”, we chose the stacked bar chart which works really nice for this type of question (Yes/No) and as can be seen in Figure 11, Figure 12 and Figure 13, it is quick to understand that for all three groups most respondents feel that society puts pressure on men in a way that is unhealthy or bad for them, even when the values or percentages for Yes/No are not shown in the chart.

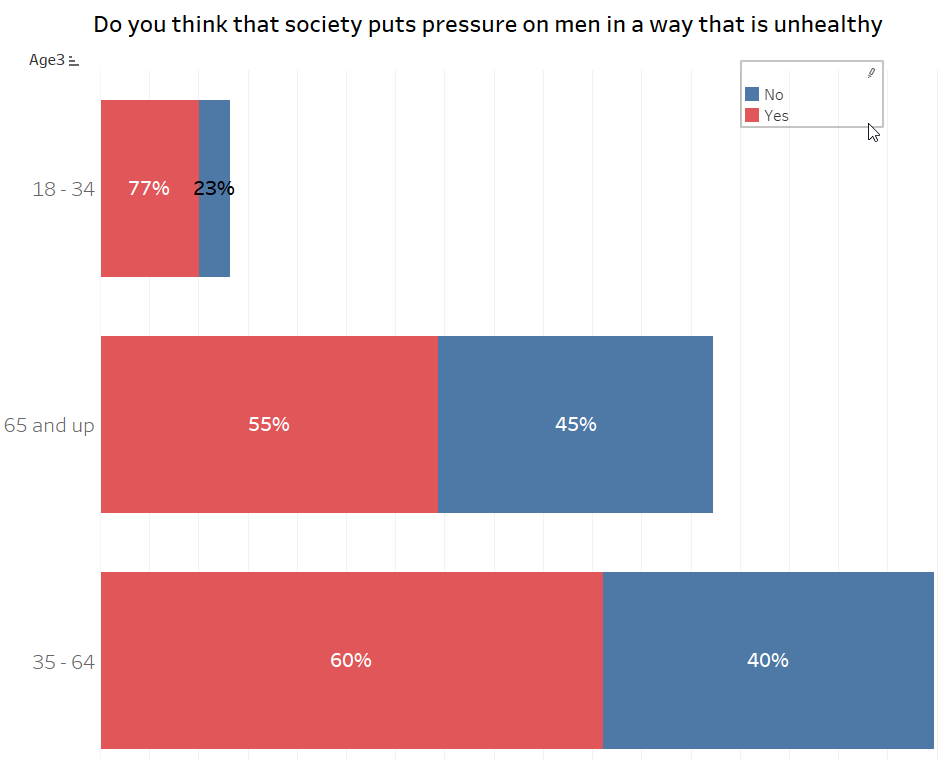


Figure 11 – Tableau

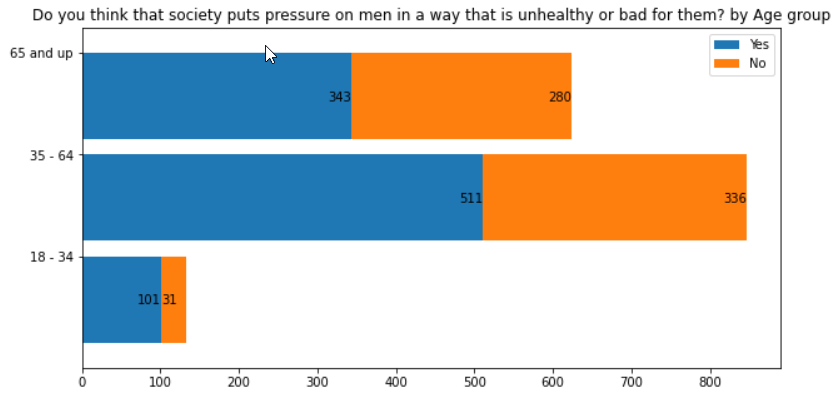


Figure 12 – Python

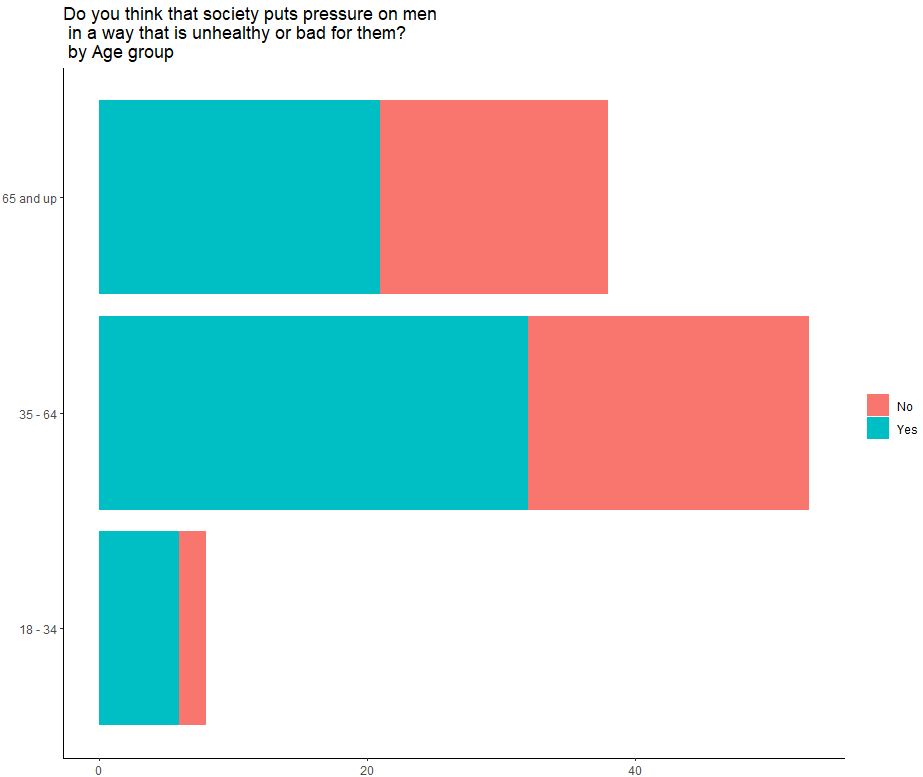


Figure 13 - R

Afterwards, to analyse the question “*How important is it to you that others see you as masculine?*” we applied different types of visualisation to be able to compare them for the same set of data. Figure 14 shows the Tableau visualisation using a line chart with the shaded area and light colours. Figure 15 shows the Python visualization using a pie chart, this question has only 4 possible answers, thus it is fine to use this type of chart. Through both charts we can clearly understand that for all age groups most men answered that somewhat is important that others see them as masculine.

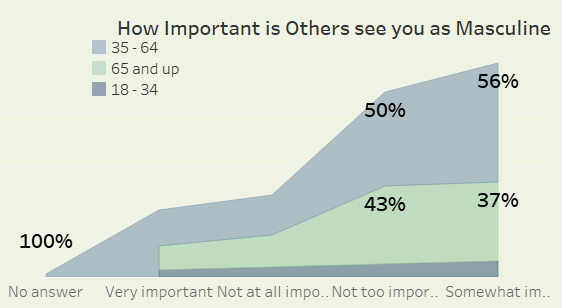


Figure 14 - Tableau

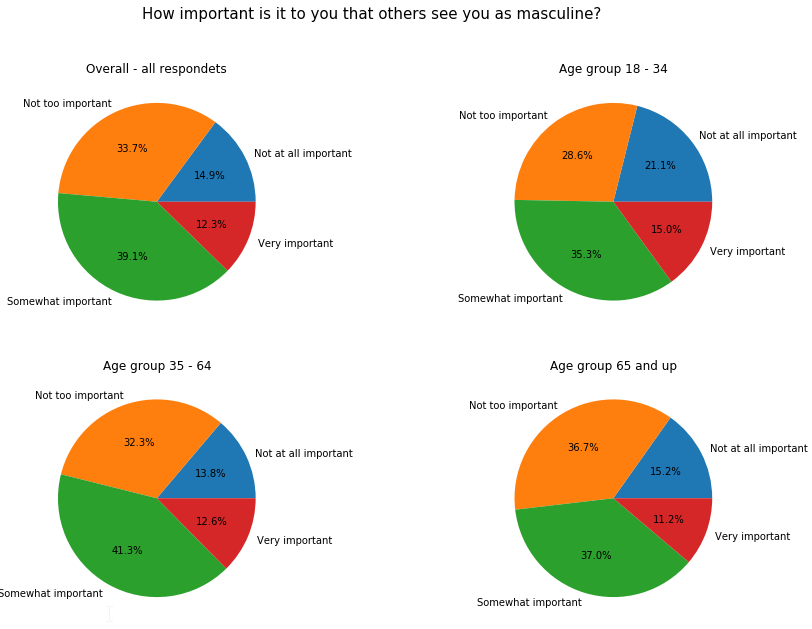
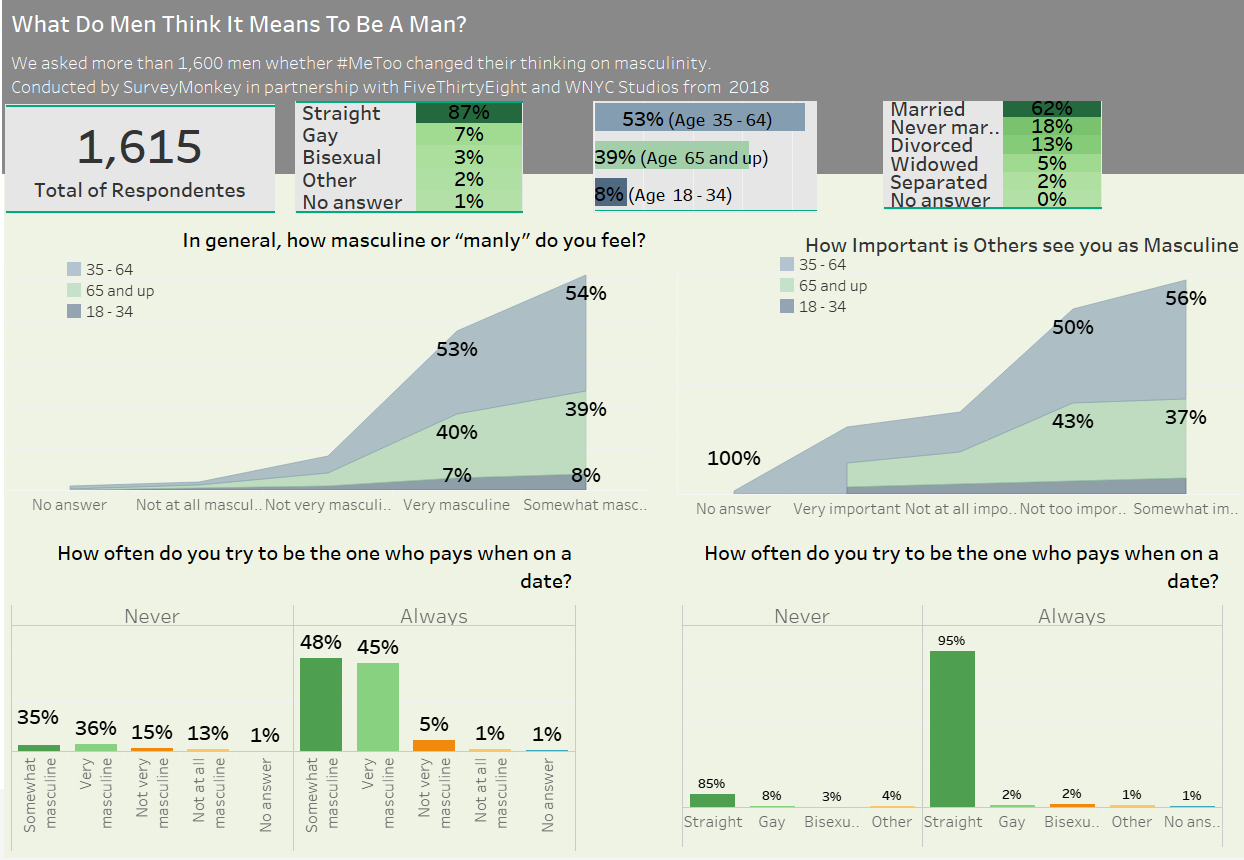


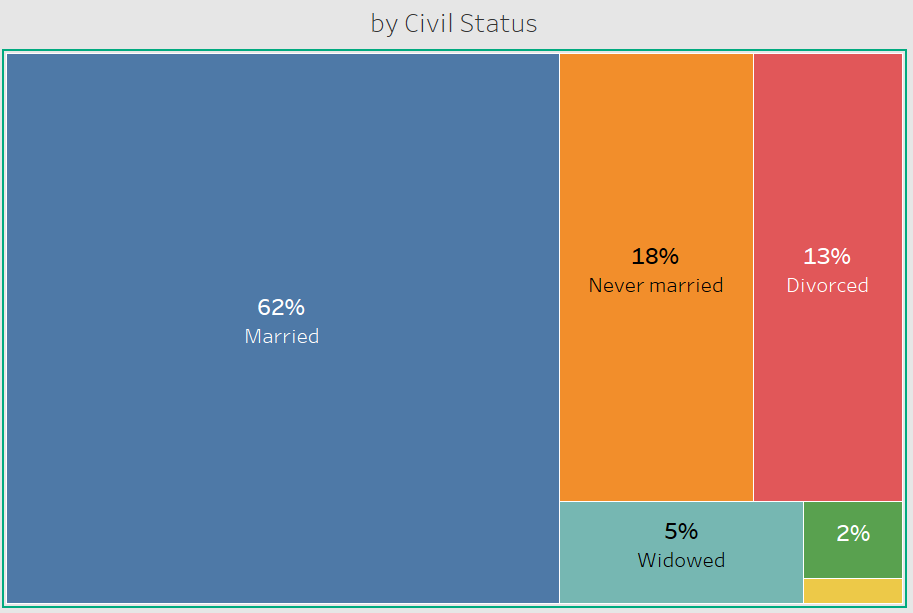
Figure 15 - Python

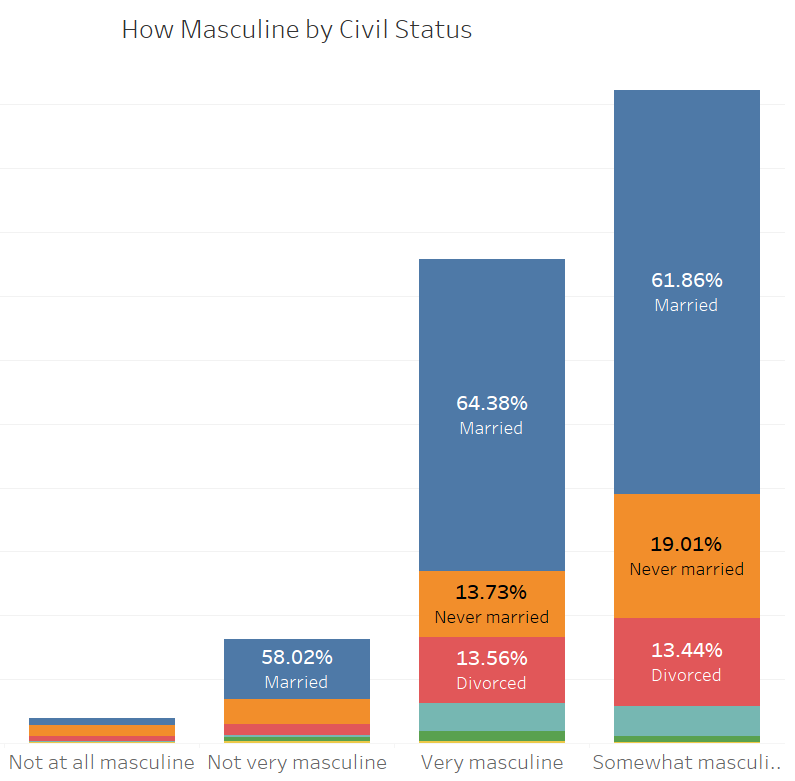
We also analysed a few other questions and distributions for which the charts can be seen at the end of this report and all follow the same approaches described above. All the three tools are very powerful, and we did not use the most advanced features for any of them due to a time issue, however, we were able to understand more about each of them in the development of this work. Thus, regardless of the tool being used, it is important to understand the data and know how to clearly represent that in a way that gets the audience attention and help them to easily comprehend what the data is showing. It is important to keep simple.

See below other visualisations developed on this task.

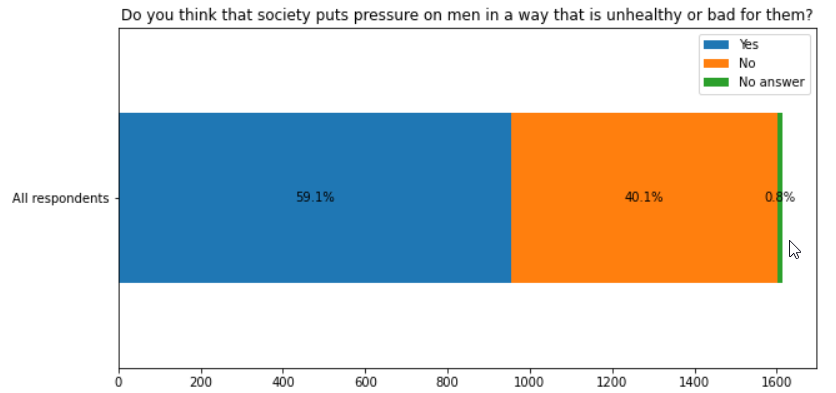
Tableau:

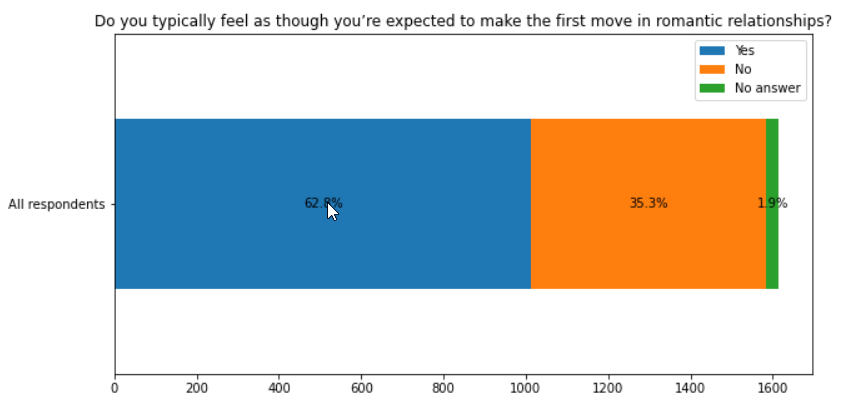


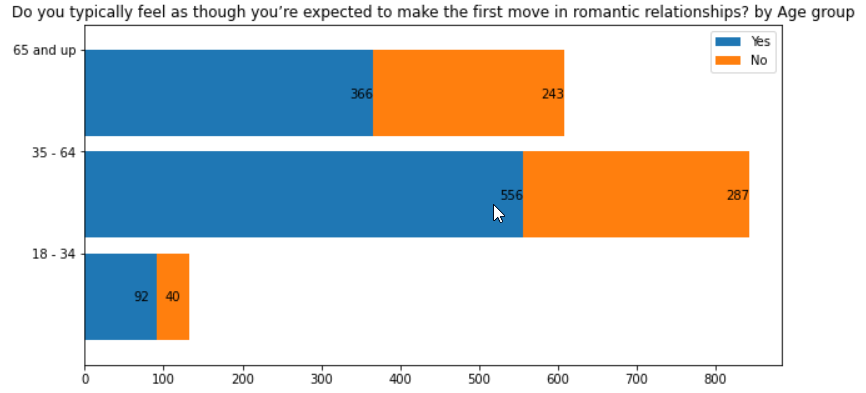


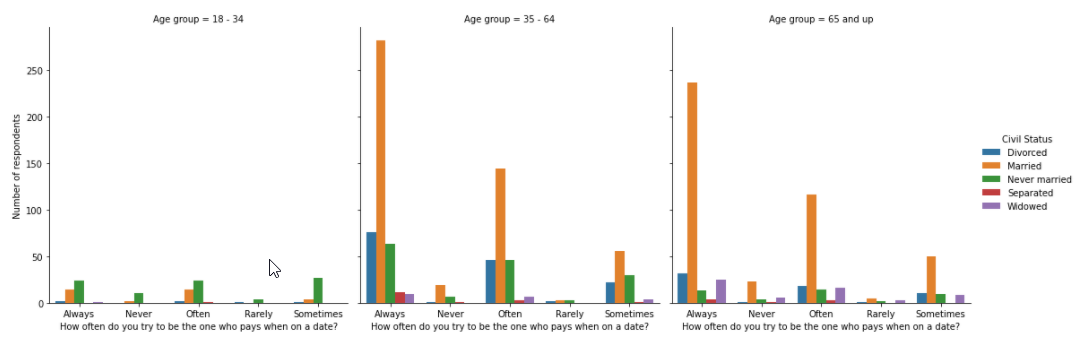


Python:

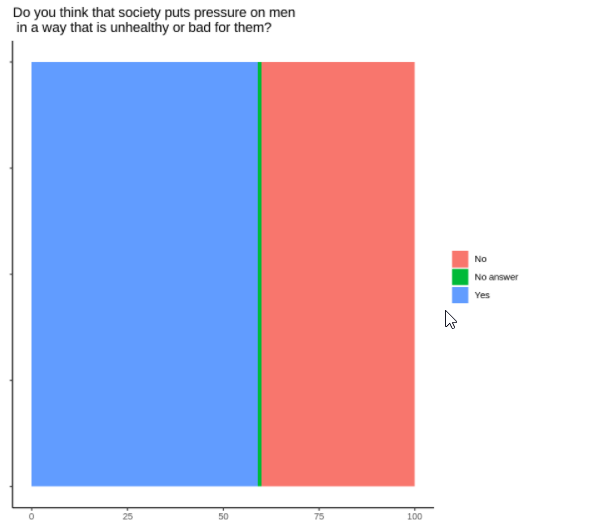


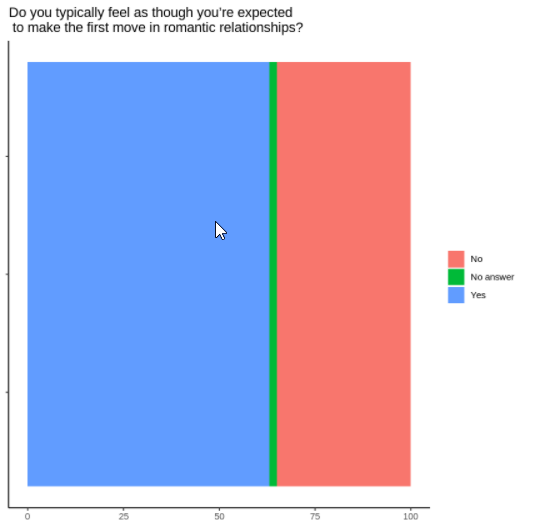


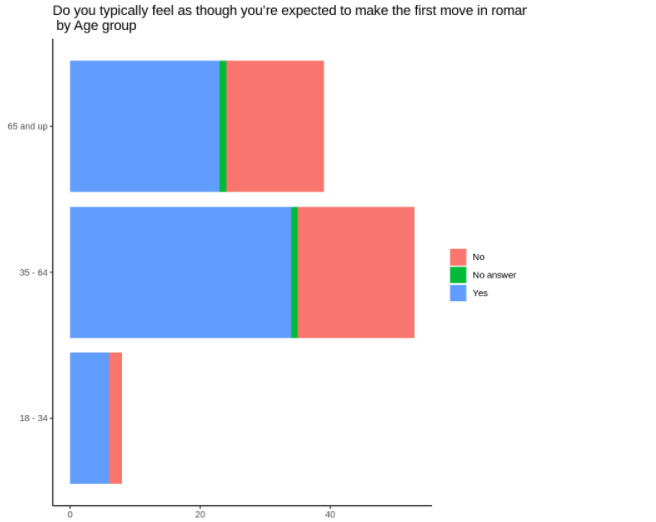




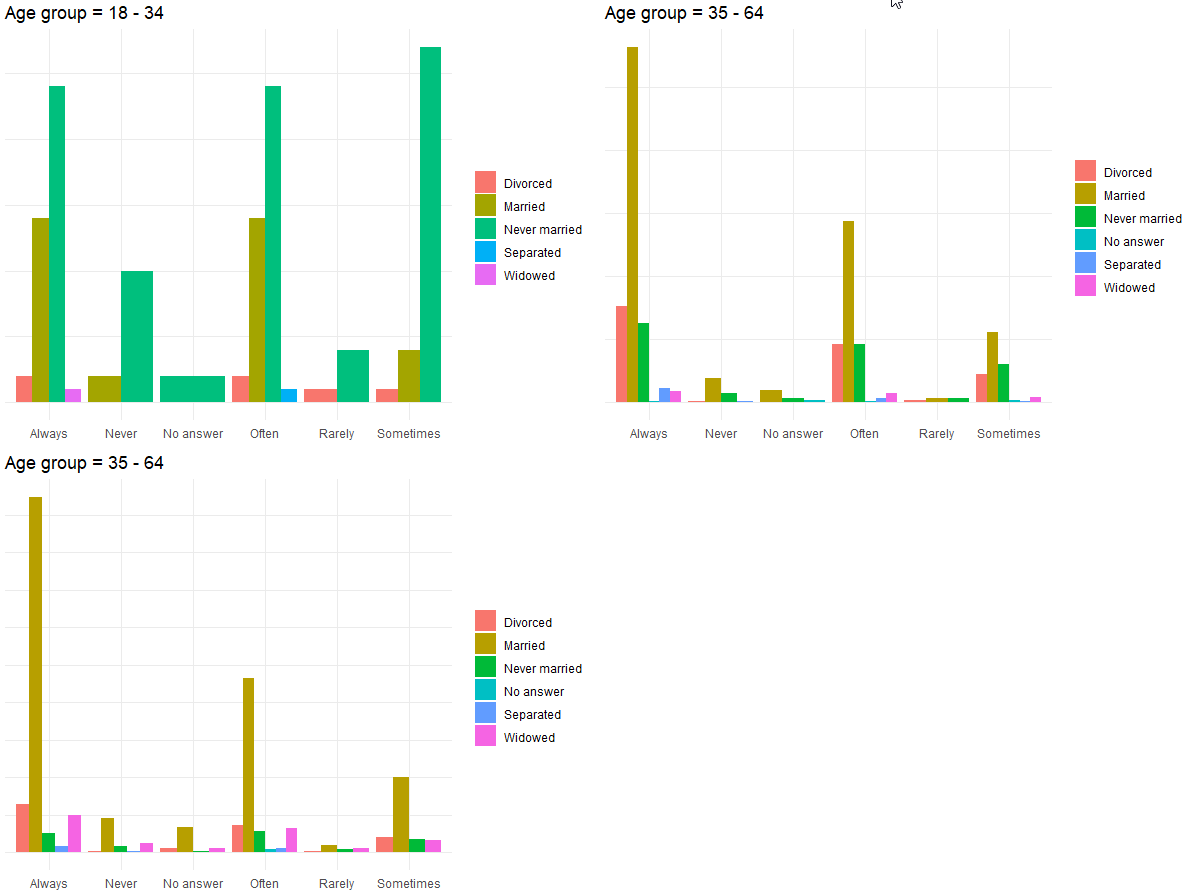
R:

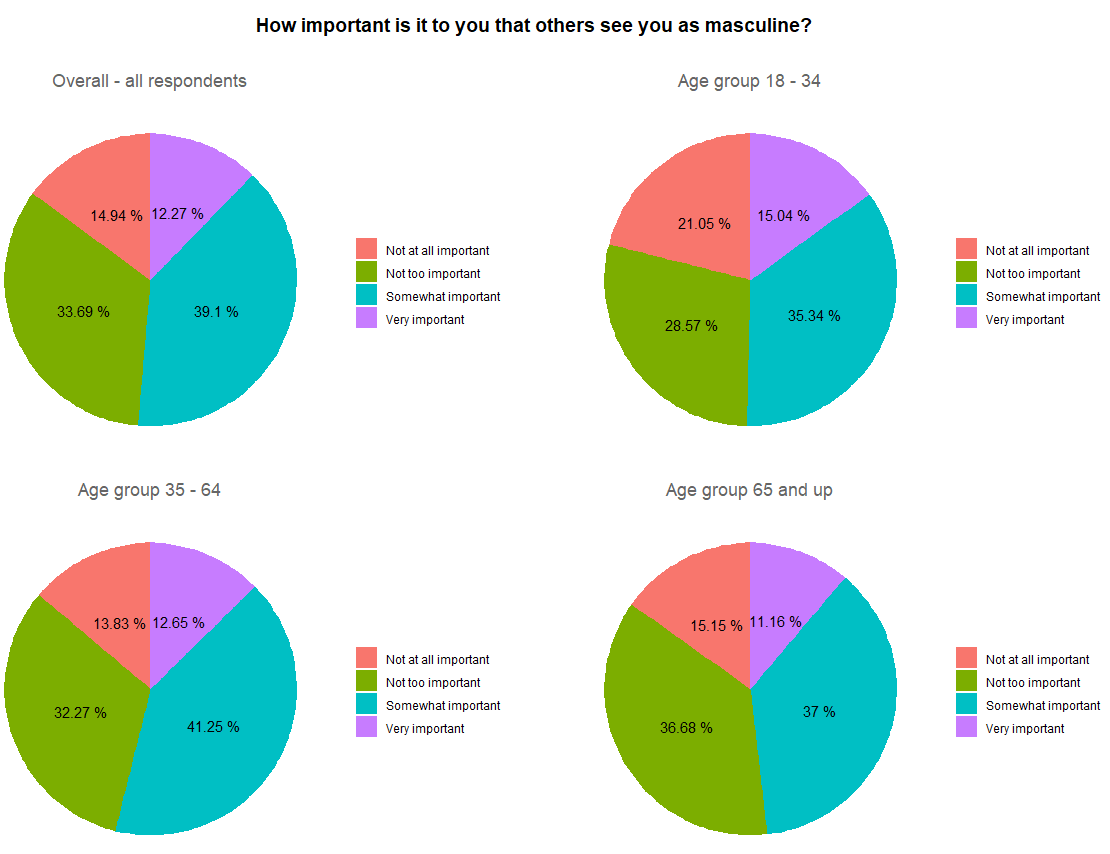






How often do you try to be the one who pays when on a date?





**References**

[1] Koeze, E. and Barry-Jester, A. (2018) ‘What Do Men Think It Means To Be A Man?’, FiveThirtyEight. Available at: <https://fivethirtyeight.com/features/what-do-men-think-it-means-to-be-a-man/> (Accessed: 17 Nov 2020)

[2] Hoare, T. (2020) ‘The Art and Science of Data Visualisation’, Dublin Business School. Available at: <https://elearning.dbs.ie/pluginfile.php/1223232/mod_resource/content/0/Tufte%20Design%20Principles.pdf> (Accessed: 17 Nov 2020)

[3] Shaffer, J. (2020) ‘Owning Your Data Story’, Adjunct Faculty, University of Cincinnati. Available at: <https://elearning.dbs.ie/pluginfile.php/1230557/mod_resource/content/0/Tableau%20Resources%20-%20Story%20Telling%20-%20Owning%20Your%20Data%20Story.pdf> (Accessed: 17 Nov 2020)

[4] Shaffer, J. (2020) ‘Pie Charts and Other Chart Types’, Tableau, Adjunct Faculty, University of Cincinnati. Available at: <https://elearning.dbs.ie/pluginfile.php/1230597/mod_resource/content/0/Tableau%20Resources%20-%20Chart%20Types%20%28Jeffery%20Shaffer%29.pdf> (Accessed: 17 Nov 2020)

[5] Hardin, M., Hom, D., Perez, R. and Williams, L. (2020) ‘Which chart or graph is right for you?’, Tableau. Available at: <https://elearning.dbs.ie/pluginfile.php/1230592/mod_resource/content/0/Tableau%20RESOURCE%20-%20Which%20Chart%20or%20Graph%20is%20Right%20for%20You.pdf> (Accessed: 17 Nov 2020)

[6] ‘Python vs R for Data Science: And the winner is..’ (2018), Data-Driven Science, Medium. Available at: <https://medium.com/@datadrivenscience/python-vs-r-for-data-science-and-the-winner-is-3ebb1a968197#:~:text=Python%20has%20caught%20up%20some,ggplot2%2C%20htmlwidgets%2C%20Leaflet).&text=Python%20is%20a%20powerful%2C%20versatile,of%20tasks%20in%20computer%20science.&text=Using%20more%20tools%20will%20only%20make%20you%20better%20as%20a%20data%20scientist>. (Accessed: 25 Nov 2020)

[7] ‘Data Visualization in R vs. Python’ (2019), R-bloggers. Available at: <https://www.r-bloggers.com/2019/12/data-visualization-in-r-vs-python/> (Accessed: 25 Nov 2020)

[8] Golchert, M. (2019) ‘Data Visualization in R vs. Python’, INWT Statistics. Available at: <https://www.inwt-statistics.com/read-blog/data-visualization-R-versus-python.html> (Accessed: 25 Nov 2020)

[9] Mendis, A. (2019) ‘R vs Python for Data Visualization’, KDnuggets. Available at: <https://www.kdnuggets.com/2019/03/r-vs-python-data-visualization.html> (Accessed: 25 Nov 2020).

[10] ‘What is Tableau? Uses and Applications’ (2020), Guru99. Available at: <https://www.guru99.com/what-is-tableau.html#:~:text=Tableau%20is%20a%20powerful%20and,form%20of%20dashboards%20and%20worksheets>. (Accessed: 26 Nov 2020)