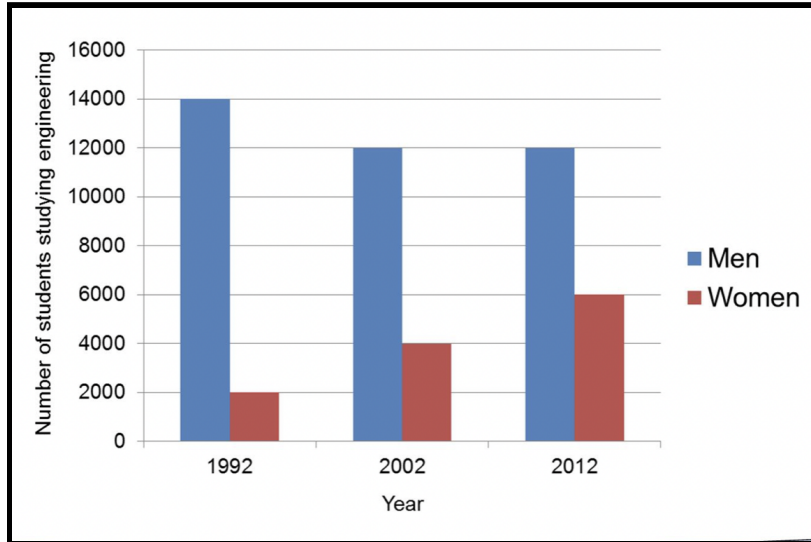


INF 143 ASSIGNMENT 1

Name: Viraj Vijaywargiya

1) Examples of visualizations

Good Visualization:

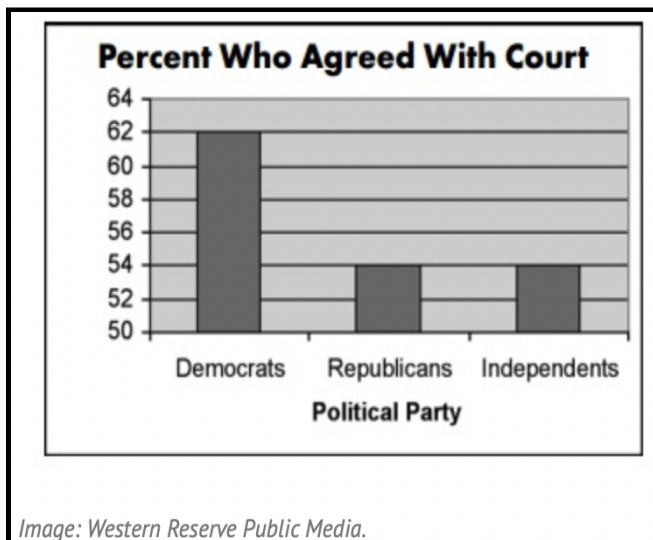


The grouped bar chart shows the number of men and women studying engineering at Australian universities between the years 1992 and 2012 with 10-year intervals. From the chart, it can be seen that the number of students, who are men, fell from 14000 to 12000 as the years progressed. Whereas, the number of students, who are women, increased from 2000 to 6000 as the years progressed. It also shows that men continue

to make up the majority of students, however, the amount of women students is increasing, and the difference between the number of men and women students is decreasing.

I consider this as a good visualization because it is a simple grouped bar chart that shows the data clearly, and it's easy to read. The type of plot, a grouped bar chart, suits the data given. The position, x and y axis, is well defined and is clear, with a good selection of variables. The different colors, blue and red, used to group men and women are easy to read and to differentiate. The visualization is labeled properly, and the ordering of data, for example Year in x axis, is natural. The bars start from 0 on the y axis which prevents misleading, and the visualization is in 2D.

Bad Visualization:



The bar chart shows the percentage of political parties who agreed with the decision to remove the feeding tube. From the chart, it can be seen that the Democrats agreed the most, having the Republicans and Independents tied.

This is a bad visualization because at first glance, it looks like almost three times as many democrats supported the decision.

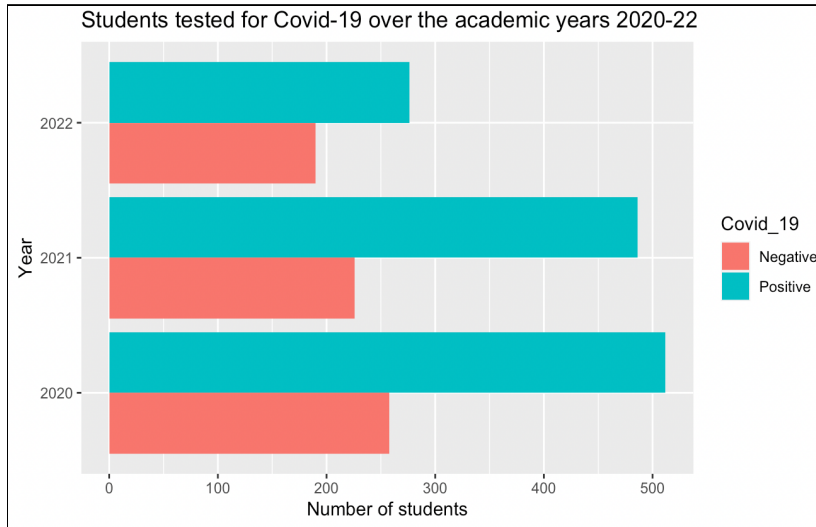
Image: Western Reserve Public Media.

INF 143 ASSIGNMENT 1

Name: Viraj Vijaywargiya

However, looking at the scale on y axis closely shows that the percentage difference between democrats and republicans is only 8%. This is because the baseline on the y axis scale should be 0, and the chart above starts at 50. This can be very misleading.

2) The creation is based on hypothetical data and not an actual one.



The visualization is a grouped bar chart that shows the number of students in a university tested either positive or negative for covid-19 from the academic year 2020 to 2022. As it can be seen from the chat, 2020 had the most positive cases for covid-19 and it decreases over the years. The chart also shows that while 2020 and 2021 have almost the same number of positive cases, positive cases in 2022

is decreased by almost half from 2020. This also shows a decrease in the amount of covid tests students took over the academic years.

This is a good visualization as it is a simple grouped bar chart that shows the data clearly and is easy to understand. The plot is labeled well with good selection of variables for grouped data and the axes. The different colors used to differentiate negative and positive covid cases is easy to differentiate when reading the graph and the ordering for both the axes is natural. In addition, the visualization is 2D and the bars start from baseline 0.

Code:

```
myData <- data.frame(Covid_19=c("Positive", "Negative"),
                     Year=c("2020", "2020", "2021", "2021", "2022", "2022"),
                     Freq=c(512, 258, 486, 226, 276, 190))

ggplot(myData, aes(x = Year, y = Freq, fill = Covid_19)) +
  geom_bar(position = "dodge", stat = "identity") +
  coord_flip() + labs(title = "Students tested for Covid-19 over the academic years 2020-22",
                     x = "Year", y = "Number of students")
```

First 3 lines: Using **data.frame** is how you create a data set, and assign it to an object, **myData**, containing the data set. The data set has 3 variables, and their values, defined in **data.frame** – Covid_19, Year, and Freq.

4th line: ggplot function plots the data. Takes in the data set, **myData**, and aesthetics - setting the **x** axis to **Year**, **y** axis to **Freq**, and **fill** (colors in the bar) to **Covid_19**.

INF 143 ASSIGNMENT 1

Name: Viraj Vijaywargiya

5th line: **geom_bar** function defines that the plot is a bar chart, setting **position** to **dodge** defines that it is a grouped bar chart, and setting **stat** to **identity** is the default stat type which leaves the data as it is.

6th line: **coord_flip** function flips the x and y axis, and **labs** function is used for modifying axis, legend, and plot labels.