DataVizA Tutorial: Introduction and R: Solutions

Department of Econometrics and Business Statistics, Monash University

Tutorial 1

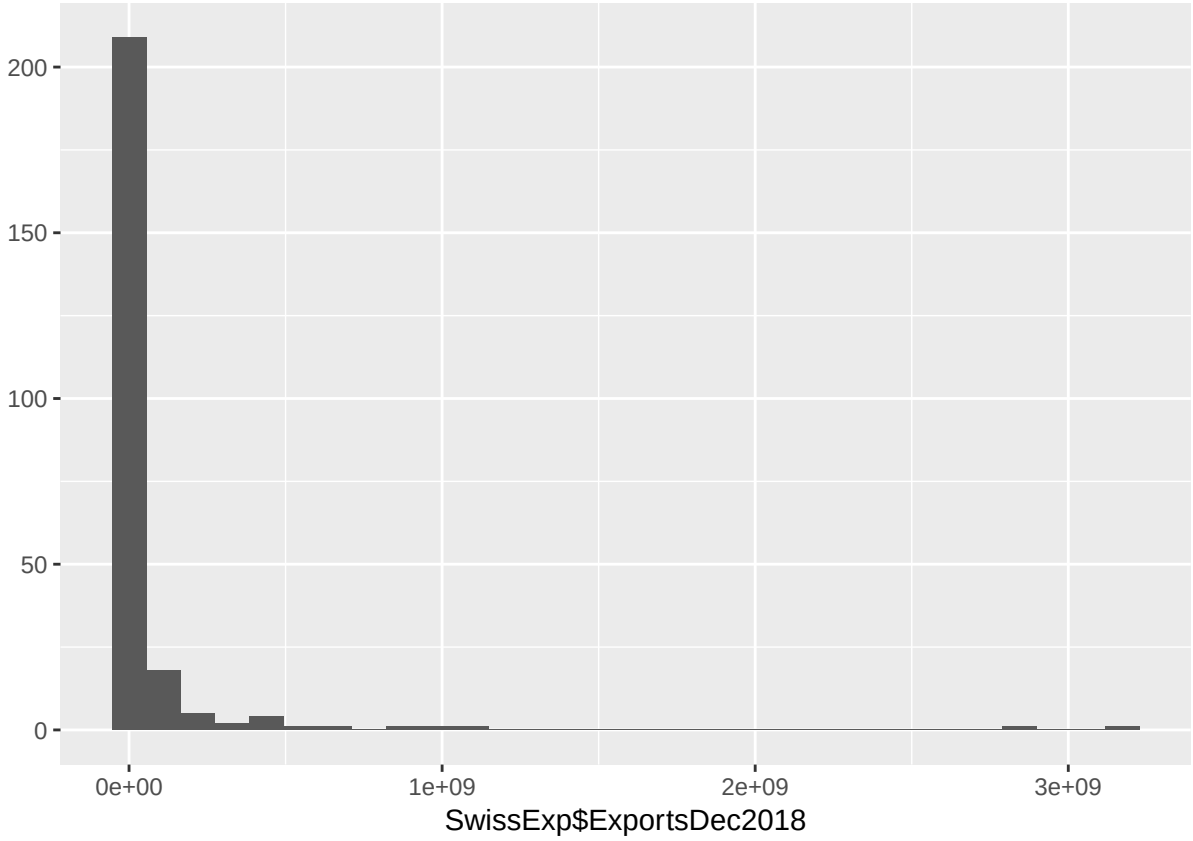
Introduction to R

1. Get R and R Studio installed on your computer.
2. Install the `tidyverse` onto your computer.
3. Load the dataset SwissExports.rds which can be downloaded from Moodle. This data has four variables, a country code, the total value of exports in January 1988 and December 2018 (both in Swiss Francs) and a region code. The data can be read using the `readRDS` function

```
SwissExp<-readRDS('SwissExport.rds')
```

4. Produce a histogram of the exports in December 2018. Use the function `qplot` for this.

```
library(tidyverse)
qplot(SwissExp$ExportsDec2018)
```



5. How would you describe the skew in the data

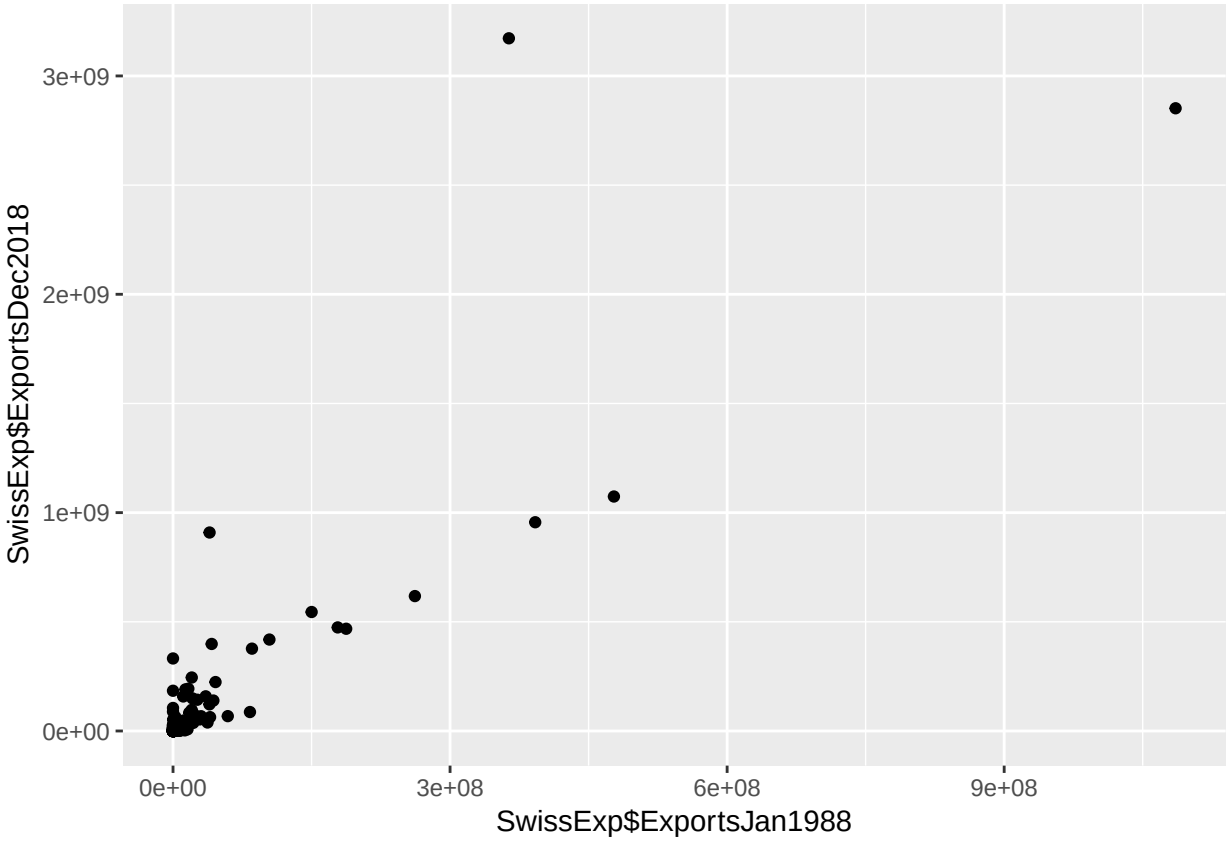
There is substantial **right** skew.

6. Do you identify any outliers in the data?

There are some outliers in the right tail, these are Switzerland's largest export markets.

7. Produce a scatterplot of January 1988 exports against December 2018 exports. Use the `qplot` function.

```
qplot(SwissExp$ExportsJan1988,SwissExp$ExportsDec2018)
```



8. Describe the relationship between exports in January 1988 and December 2018.

In general there is a positive relationship as indicated by the upward sloping pattern (from bottom left to top right) in the scatterplot. This suggests stability in most trading relationships.

8. Are there outliers in the data?

There are two outliers in the data. Further inspection shows that these are Germany and the USA. In the case of Germany, it was by far Switzerlands largest export market in 1988 and is still the second largest market. The US was the largest export market in December 2018 but was only the fourth largest in January 1988.

9. How could this plot be improved?

Most countries are grouped too closely into the bottom left hand corner. A log-log scale would help to visualise these better. Also, points could be replaced with country codes.