DataVizA Tutorial: DataMunging

Department of Econometrics and Business Statistics, Monash University

Tutorial 6

First Normal Form

1. Discuss whether the following databases satisfy first normal form.

Database A:

Name	Social Media Username
Jane Smith	Facebook: jsChampion
Kamaru Usman	Twitter: kusman, LinkedIn: lx99
Li Xiao	WeChat: lx99

Database B:

Name	Social Media Username
Jane Smith	Facebook: jsChampion
Kamaru Usman	Twitter: kusman
Kamaru Usman	LinkedIn: lx99
Li Xiao	WeChat: lx99

Swiss Exports: Full Data

The file *SwissExportsFull.csv* contains the full export data for Switzerland. Each row represents a different date. The first column is the date variable, the second column is the year only and each remaining column measures exports to a different country.

2. Read the data into R
3. Get the data into long form using the `gather` function
4. Recall that in the previous tutorial, one issue was that monthly data were noisy. Using `group_by` and `summarise` create a new dataset of yearly aggregate exports to each country.
5. Plot time series line plots of Swiss exports to Germany (DE), the USA (US), China (CN) and India (IN). Facet by country.
6. Plot these four lines on a single plot with each country in a different colour. Hint: Use the aesthetic
7. Comment on these plots
8. Now produce a scatterplot on a log-log scale of 1988 exports against 2018 exports. Use country abbreviations rather than points
9. Produce the same plot but remove all countries for which exports are zero in either 1988 or 2018.