

# DataVizA Tutorial: Plotting many variables

*Department of Econometrics and Business Statistics, Monash University*

## *Tutorial 5*

### Swiss Data

1. Repeat question 3 from last week but use text to represent the country and color to represent the region. Note these regions are very loosely defined for instance Mexico could easily be in North America, and the Central Asian (CA) could be better named the former Soviet region, etc.
2. Discuss this plot.
3. Since the previous plots suffer from a little overplotting, repeat the analysis but facet by region.

### U.S. States Data

The data in the file *USStateData.rds* contains information on demography and geography for the states of the United States. These data were measured in 1977.

4. Construct a bubble chart of income per capita (x axis) against life expectancy (y axis). The size of the bubble should reflect the State population, and the color should reflect the region.
5. Discuss this plot.
6. Add the state abbreviations as text on the plot
7. If you are not satisfied with your answer to question 6, redo the question. Hint: you may need to specify some aesthetics within a geom.
8. The file *USStateRed.rds* contains the same data but with the variables **State**, **Abbreviation**, **Frost**, **Murder** and **HS Grad** stripped out. Load this data and plot a ggpairs plot.
9. Discuss some interesting features of the plot.