

# Homework Exercises Information Visualisation

A. Posekany

May 9, 2017

## Exercises

For each of the following exercises implement a R shiny app which visualises and illustrates the most important descriptive facts of each data set variable and the data set as a whole.

1. Explore the variables "Fertility", "Agriculture", "Education", "Catholic" and "Infant. Mortality" of the data set *swiss* in the R package *utils*. Consider each variable as a separate sample. (Look up what the variables mean for interpretation with `?swiss`.) For each variable:
  - Consider reasonable estimates of location, scale, skewness and peakedness.
  - Use appropriate graphical tools to visualise the data and explain the observed behaviour.
  - Can the data be interpreted to be (approximately) normal? Do the data contain outliers? (Robust estimators? Consequences for inferential methods that could be applied?)

What can you say about the relations between the variables? (Look at a scatterplot matrix.)

2. Explore the variables "Population", "Income", "Illiteracy", "Life.Exp" and "Murder" of the data set *states77*. For each variable proceed as before:
  - Consider reasonable estimates of location, scale, skewness and peakedness.
  - Use appropriate graphical tools to visualise the data and explain the observed behaviour.
  - Can the data be interpreted to be (approximately) normal? Do the data contain outliers? (Robust estimators? Consequences for inferential methods that could be applied?)

What can you say about the relations between the variables? (Look at a scatterplot matrix.)

3. Explore the data set *LakeHuron* with and without taking into account the time series aspect.
  - Consider reasonable estimates of location, scale, skewness and peakedness.
  - Use appropriate graphical tools to visualise the data and explain the observed behaviour.
  - Can the data be interpreted to be (approximately) normal? Do the data contain outliers? (Robust estimators? Consequences for inferential methods that could be applied?)

4. Explore the data set Titanic. How do age/sex/travel class influence survival? Is there an indication that Simpson's paradox is masking some marginal results?
5. Model the variable Education of the data set swiss in the R package *utils* employing all other variables.
  - Consider the assumptions for linear regression. Which variables may be included in such an analysis a priori?
  - Use appropriate graphical tools to visualise the quality plots for regression output and check all assumptions for regression visually.
  - Are there leverage points? Which observations should be used in a final model? Are transformations of these variables required or advisable?
  - Write a the initial regression model and the finally selected regression estimate.
6. Model the variable "Murder" of the data set states77 employing all other variables.
  - Consider the assumptions for linear regression. Which variables may be included in such an analysis a priori?
  - Use appropriate graphical tools to visualise the quality plots for regression output and check all assumptions for regression visually.
  - Are there leverage points? Which observations should be used in a final model? Are transformations of these variables required or advisable?
  - Write a the initial regression model and the finally selected regression estimate.
  - Comment on former president G. Bush' statement: "Poor people are not necessarily killers."

Project: **"Educational App for high school"**

From a selected choice of high school topics choose a good way to interactively let students learn and explore the influence of choice of variables in mathematical contexts and functions.

Consider visual attractiveness for a young audience as well as practicality and that it should be clearly focussed not distracting students with an overload of information. Be concise!