Homework 1

For all questions involving histograms, choose a sensible binwidth and breakpoints, unless otherwise indicated.

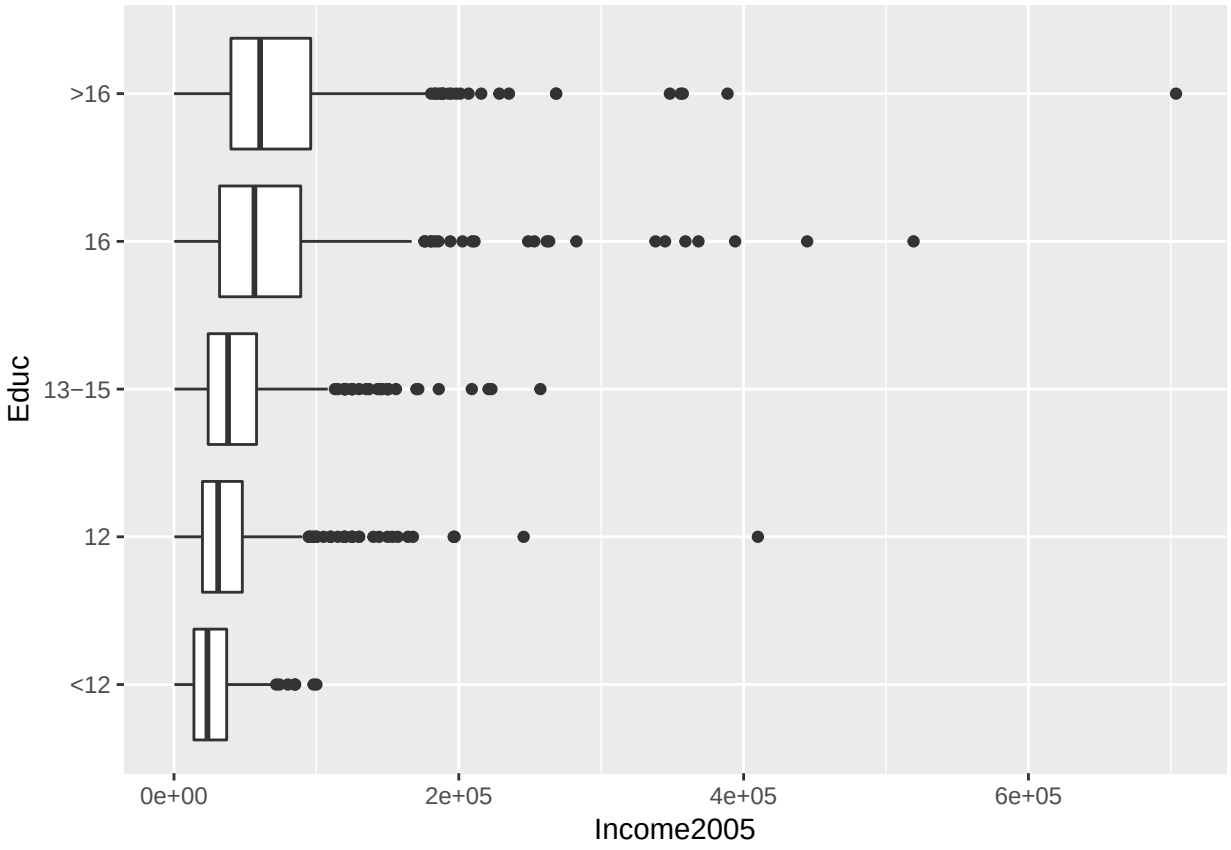
1. Income

a) Describe in detail the features you observe in the boxplots below, plotted with data from the *ex0525* dataset, **Sleuth3** page. (see page 29 in *Graphical Data Analysis in R* for a list of features to concentrate on, and the numbered list on the bottom of page 43 for an example of how to describe features of a graph in words.) [5 points]

```
#install.packages("Sleuth3")
library(Sleuth3)
library(tidyverse)

# convert Educ from an integer to a factor, and make "<12" the first factor level
mydata <- ex0525 %>%
  dplyr::mutate(Educ = forcats::fct_relevel(Educ, "<12"))

ggplot(mydata, aes(Educ, Income2005)) +
  geom_boxplot() +
  coord_flip()   # for horizontal boxplots
```



b) Plot a histogram of the `Income2005` variable in the dataset referenced in part a). [3 points]

c) Use `+facet_wrap(~Educ)` to facet the histogram on education level. [3 points]

d) What do you learn from the histograms that wasn't apparent in the boxplots from question 1? [3 points]

2. Respiratory Rates

a) Plot right closed and right open histograms for each of the two variables in the *ex0824* dataset in the **Sleuth3** package using default binwidths and breaks. (4 histograms in total). [4 points]

b) For which variable, `Age` or `Rate`, do the two versions differ more? Why? [3 points]

c) Redraw the `Age` histograms with different parameters so that the right closed and right open versions are identical. [3 points]

3. Movie budgets

Are there rounding patterns in the `budget` variable of the *movies* in the **ggplot2movies** package? If so, what are the patterns? (Note: according to the textbook this dataset is in the **ggplot2** package, but it has since been moved to a separate package.) Support your conclusions with graphical evidence. You are encouraged to break the variable down into different budget ranges and consider them separately. [8 points]

4. Finches

a) Plot separate density histograms of the beak depth of the finches in *case0201* from the **Sleuth3** package, with density curves overlaid as on page 34 of the textbook. (However, do this by facetting on `Year` rather than using `grid.arrange`). [3 points]

b) Plot both density curves on the same graph to facilitate comparison. Make 1976 yellow and 1978 blue. Use alpha blending so the fills are transparent. [3 points]

c) Based on your graphs in parts a) and b), describe how the distributions differ by year. [3 points]

d) What is the cause of the difference according to the information in the help file? [3 points]

5. Salary

Is the `Salary` variable in the *case0102* of **Sleuth3** normally distributed? Use two different graphical methods to provide evidence. [6 points]