seit 1558

Friedrich-Schiller-Universität Jena

**Wirtschaftswissenschaftliche Fakultät**

Lehrstuhl für Empirische
und Experimentelle
Wirtschaftsforschung
**Prof. Dr. Oliver Kirchkamp**
Saturday, 11 August 2018

Carl-Zeiss-Str. 3
07737 Jena
oliver@kirchkamp.de
http://www.kirchkamp.de

**Exam "Using Graphs and Visualising Data"**

You can use the handout and any documentation for software that we have discussed in the lecture. If during the exam you have questions regarding the organisation of the exam, please call 03641-7962943.

You will submit your answers as a text file with your R commands to `oliver@kirchkamp.de` before 10:15.

Your email will contain no other files. Your email will contain no output from your commands. Your email will contain no attachments in PDF or Microsoft Word format.

In the following you will have a look at the data frame `Benefits` from the library `Ecdat`.

1. In this data frame the variable `rr` contains the "replacement rate", i.e. the ratio of the benefits of an unemployed person to the last wage of this person.

   Generate an empirical cumulative distribution function plot (ecdfplot) of `rr`.

2. The variable `sex` tells you whether the unemployed person is male or female. Generate one graph with two ecdfplots: one for males and one for females. This graph should allow you to easily compare the two groups.

3. Which other options do you have to produce two ecdfplots of `rr` in one graph. Compare these options and add a comment with your answer to your `R` file.

4. You suspect that `age` might be a variable which has an impact on `rr`. Provide a graph which allows to investigate this hypothesis. Compare in a comment your graph with other types of graphs. Explain why you have chosen this type of graph.

5. Does `tenure` have an impact on `rr`?

   a) Provide a graph which allows to answer this question?

   b) Does this answer depend on whether the unemployed person is `male` or `female`? Provide a graph!

6. Generate a scatterplot of the joint distribution of `age` and `yrdispl`.

   a) One problem with this scatterplot is that you have multiple observations with identical values. Explain why this is a problem!

   b) Compare at least three different options to show the joint distribution of `age` and `yrdispl` which do not suffer from this problem. For each option show a graph.

   c) In which way does the joint distribution of `age` and `yrdispl` depend on `sex`?