DAT 301 Written Assignment 1

(due 11:59pm Friday, 09/23/2022)

Install Latex Distribution

One of the goals of this lab is that you make sure you can create interactive document files (Rmd) and convert them to pdf, as well as html format. In order to be able to write mathematical text, you would need Latex, wich RStudio uses to produce such text in pdf files.

Latex is a free typesetting system/software that for quite some time has been a standard for writing books, scientific articles, theses and other documments in mathematics, statistics, physics, computer science and other disciplines.

If you've never used Latex on your machine, you will need some Latex distribution (MikTex, MacTex, TinyTex,...), which is a set of packages required for Latex. You can find various online tutorials how to install them. For example, MacTex.

You can read instructions here: https://www.latex-project.org/get/. Or else,

- for Windows   • for Mac   • for Linux/Ubuntu

Once you are done with Tex installation, you can (but don't have to) also install some Latex editor, if you want to create math/stats pdf files independently of RStudio, i.e. for some purposes outside of this class. There are various Latex editors. I use TexShop.

Hopefully, your MacTex installation went smoothly. Make sure you can create a pdf file using "Knit" button in RStudio. To do that, you need to create a new Rmarkdown file (`File --> New File --> R Markdown --> Document --> PDF`). Make sure this works or you fix problems ASAP. Let me know if you have any trouble ASAP. Don't wait couple of days before the lab deadline. You won't be able to finish the lab.

The above sequence of clicks should produce an Rmd file, which you can see it open in your RStudio. It is a template with some content that helps you figure out how basics of RMarkdown work. Save this file in order to run it, by clicking on `save` button (you would need to choose a name of the file). To run the file, click on `Knit`. Since you chose to create PDF (rather than HTML or Word), it will automatically generate pdf file.

Absenteeism at Work

The file `Absenteeism_at_work.csv` is posted on Canvas and course website, as well as the documentation about the data. This is taken from UC Irvine Machine Learning Repository archive https://archive.ics.uci.edu/ml/datasets

Sample RMD

A sample RMD file has been posted here for your reference.

Instructions

- Read the documentation to become familiar with the meanings of the variables/columns.

- Read in the data set using the command

```
df = read.csv("Absenteeism_at_work.csv",sep=";",header=TRUE)
```

- You will onle need to submit one PDF file, produced by your Rmd file. Include your code, plot and comments in your Rmarkdown file, so that they are shown in the pdf file.

- In each plot, include appropriate title and labels. Include the legend, if appropriate. Also, after each plot, write a short comment (one or two sentence) if you see something on the graph, i.e. if graph reveals or suggests something about the data. **Do not forget** to write these comments, even if you can't say much by looking at the graph (in that case, just say that the graph is not very useful, i.e. doesn't suggest anything).

- Use base plot (this time), not `ggplot2`.

1. Plot the scatter plot of height vs. weight (so, weight on x-axis) including all the (non-missing) data.

2. Plot the histogram of hours of absences. Do not group by ID, just treat each absence as one observation.

3. Plot the histogram of age of a person corresponding to each absence. Do not group by ID, just treat each absence as one observation.

4. Plot the bar plot of hours by month. So, each month is represented by one bar, whose height is the total number of absent hours of that month.

5. Plot the box plots of hours by social smoker variable. So, you will have two box plots in one figure. Use the legend, labels, title. Play with colors.

6. Plot the box plots of hours by social drinker variable. So, you will have two box plots in one figure. Use the legend, labels, title. Play with colors.

When you are done, before submitting, go over the instructions again and for each plot make sure you met all the requirements.