Homework assignment #4

Milica Cudina

2021-09-14

Cafe data

(2 points) First, you should read the data from our csv file "cafedata.csv" into a data.frame called `cafe.data`:

If you want to see what your data.frame looks like, you can click on it in the **Global environment** in the upper right pane. The data.frame will get displayed in the upper left pane.

(2 points) Now, you interested in the types and names of the variables in your data.frame. What do you run?

You see that the days/**cases** all have corresponding **rows**. They are labeled by the row indices. The **column** names stand for the variable names.

Then, you can do a bit of exploratory analysis.

Problem 1. (3 points)

Plot the histogram of daily sales amounts. Make sure that your plot has the main title and that the axes are also labeled.

Problem 2. (3 points)

Draw the **time plot** on how the daily sales amounts evolved with time. Make sure that you connect the dots in your plot. Make sure that your plot has the main title and that the axes are also labeled.

Problem 3. (2 points)

What is the maximum number of wraps sold on a Tuesday? Use `R` commands, please, do not just look it up in the spreadsheet - that defies the purpose! Be careful about the **type** of data your number of wraps is recorded as. *Hint:* Typing `?numeric` in the RStudio console might help.

Problem 4. (3 points)

Create a stacked dot plot of the number of sandwiches sold daily. Be careful about the **type** of data your number of sandwiches is recorded as. *Hint:* Typing `?numeric` in the RStudio console might help.

Problem 5. (5 points)

What else could we ask? Think of a question the cafe manager might be interested in. Identify the relevant variable(s) from `cafe.data`, chose an appropriate plot, plot it, and comment on what you notice.

Problems

Problem 6. (15 points)

Source: Problem 2.1.5. from Pitman.

Given that there are 12 heads in 20 independent tosses, find the probability that at least two of the first five tosses landed heads.

Note: You cannot solve this kind of a problem simply using `R`*. You have to use analytic methods. It is OK to leave binomial coefficients in your answer; simplify the remainder of your expression as much as you can.*

Problem 7. (5 points)

A pair of dice is thrown. Find the probability that the sum of the outcomes is 10 or greater **if** a 5 appears on the first die.

Problem 8. (5 points)

Remember that The University of Texas has a subscription to *The Economist* and that you can study and enjoy their charts and articles whenever you want.
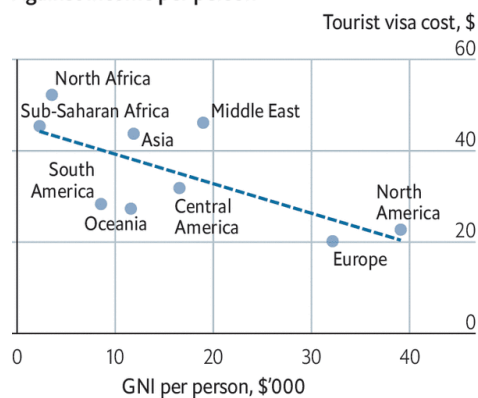
Consider the following charts and choose which of the offered statements is/are **correct**. **Justify your response!** Even if correct, answers without a justification will receive zero credit.

```
knitr::include_graphics("visas.png")
```
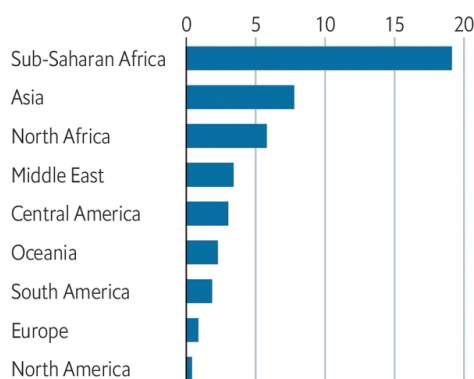

Send me your better-off...
Tourist visa costs, by region

Source: "Assessing Visa Costs on a Global Scale", by E. Recchi et al., EUI Working Paper, 2020

The Economist

a. The shortest number of work days needed to purchase a visa is for Europe.
b. The more affluent the region, the longer one needs to work to purchase the visa.
c. One has to work more than three times as long in Sub-Saharan Africa than in the Middle East to purchase a visa.
d. One has to work longer in Oceania than in Asia to purchase a visa.
e. None of the above statements are correct.

Problem 9. (5 points)

Two independent coins are tossed. If E_1 is the event "heads on first coin", E_2 the event "head on the second coin", and E_3 the event "the coins match; both are heads or tails". Which of the following statements is/are **not** true? **Justify your response!** Even if correct, answers without a justification will receive zero credit.

a. E_1 and E_2 are independent.
b. E_2 and E_3 are independent.
c. E_1 and E_3 are independent.
d. E_3 and E_1 are independent.
e. E_1, E_2 and E_3 are independent.