

In groups of 3, please assign a letter to each group member and whenever it's your turn, explain a potential solution to the following tasks to your peers. Feel free to ask me anytime to settle a dispute. Of course you can also check your notes, the slides, the internet or the tea leaves at the bottom of your cup.

1. A. If you have a dataset like the one below: how do you read it into R?

Measurements of reed (German: Schilf) at the beautiful lake "Fantasea" in 2015.
Conducted by Arno Nym, with the help of Mr. T and the B-team, as well as K. rider.

| Site | moisture | plant height | stem_diameter |
|------|----------------|-----------------|---------------|
| H1 | wet | 154 | 1.8 |
| H1 | moist | 120 | 0.95 |
| H3 | moist | 173 | 2.6 |
| H1 | dry | 134 | 1.25 |
| K7 | forgot to log, | sorry professor | 143 1.35 |
| K2 | wet | 167 | 2.7 |
| ... | | | |

2. B. Which 3 (or more) things are bad about this file? What should `str` of the object return (approximately)?

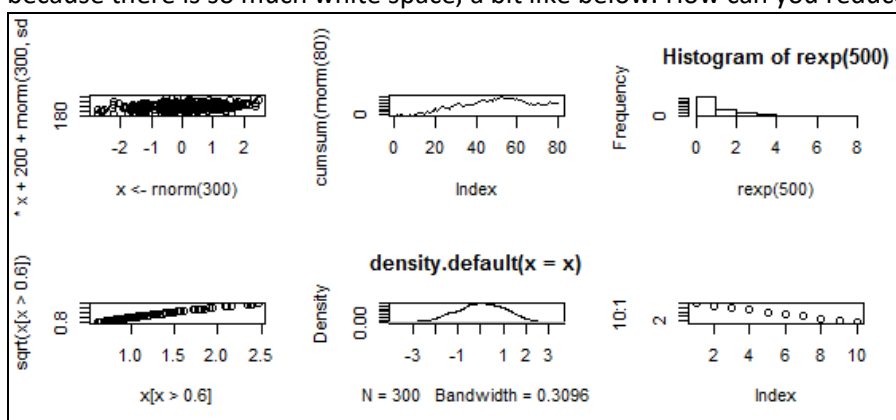
3. C. Sketch out some of the common data and object types. What does `c(pi, "pi")` return and why? Can you use the first element of that result to compute the circumference of a weed plant with 1.8 cm diameter? I meant reed plant, excuse me.

4. A. You also have a dataset from Professor Floyd (his favorite color is pink, which he never fails to mention). His PhD slaves measured salinity, dissolved organic carbon concentration and a few other chemical components you don't quite understand for each measurement site. How can you join that information with the previous table, making sure it is repeated in each observation at a site (no rows are lost, even if not available in both datasets)?

5. B. How can you create a visual representation of the data to examine the correlation between one of those weird chemical concentrations and reed plant growth?

6. C. How can you set up a 3 x 2 multipanel plot to blast away the reader of your article in "Science"? Let's assume you want to present 6 different correlations from the really important variables in the dataset for which you have already p-hacked your way to statistical significance.

7. A. Now that basic setup is not blasting anyone away, except for your dog who ate the printout and got sick, because there is so much white space, a bit like below. How can you reduce the whitespace around each figure?



8. B. Sadly, now the x and y axis labels have disappeared, and Professor Floyd starts to wonder whether you should switch PhD advisor, maybe also because your favorite color is black. "That's not even a color!". What do you program to console him?

9. C. Professor Floyd is delighted with your new graph and then remembers that old dataset with about a hundred different columns that that one PhD student was supposed to analyse but then he quit and now the data have just been lying around with nobody using it and it's just a pity how unreliable you young guys are these days and could you please make a pretty graph like that, but then one where each panel is a scatterplot of the target variable against one of those 100 columns? The point color should be pink if there are values ≥ 20 .