

- **NO** late submission will be accepted, except under special circumstances.
  - Homework must be done individually and not in groups. Discussion of problems with others is permitted (and encouraged!), but you must write your own work in your own words.
  - Submit your answers (via Canvas) as a single RMarkdown file that can be run on anyone's machine (i.e., that doesn't refer to your local files or directories). Your file name should have the following format: `lastname-NetID-week04.Rmd`. Make sure that your Rmarkdown file has yourself as author and has `output:html_document`.
  - Be sure to include detailed explanatory text and remarks of what you are doing—don't just show a lot of R code and computer generated output. Use commands from the **tidyverse** and pipes whenever you can.
1. Read R for Data Science, Chapters 14.4–14.7, 19, and 21. Do the problems as you go along (you don't need to hand them in).
  2. Download the data set <https://github.com/rfordatascience/tidytuesday/tree/master/data/2019/2019-02-19> consisting of PhDs awarded by field.
    - (a) Split the data by `major_field` and use `map()` to calculate, by `major_field`, the  $R^2$  (i.e., the coefficient of determination) for the linear model `n_phds ~ year + field`. Using `ggplot2`, make a set of boxplots of  $R^2$  by `broad_field`. Comment on what you find.
    - (b) Using `min_rank(desc())`, for each `field`, rank each year by the number of PhDs awarded. For each `major_field`, rank each field by number of PhDs awarded (aggregated over all years). For each `broad_field`, rank each year by number of PhDs awarded (aggregated over all fields).
    - (c) For each `field`, find every year with more PhDs awarded than the previous year. Find the quartiles for number of PhDs awarded by `broad_field`.
    - (d) Write a function that inputs a character string for `broad_field` and returns the signal-to-noise ratio (i.e., mean divided by standard deviation) aggregated over field and year. Also, write the function so that an error message is displayed if the character input does not belong to `broad_field`.