

- **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-week06.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 the chapter "Scraping Data" in *Data Wrangling with R* by Boehmke. It's a little dated, but still worth the read.
 2. Choose ONE of the following two websites to scrape and clean:
 - (a) The weather history for last week's class on March 4 from the closest (small) airport to Piscataway can be found at <https://www.wunderground.com/history/daily/us/nj/somerville/KSMQ/date/2019-3-4>. Download (and clean) the table at the bottom of the page under the heading "Daily Observations".
 - i. Create three separate plots (3×1 layout using `facet_wrap()`) of the time of day on the x -axis against temperature, humidity, and windspeed for the day on the y -axis (using points connected by lines for each), coloring the point by the amount of precipitation. Adjust the vertical scales accordingly.
 - (b) The weather forecast for next week from the closest (small) airport to Piscataway can be found at <https://weather.com/weather/hourbyhour/1/USNJ0524:1:US>. Download (and clean) the data using the code

```
forecast <- "https://weather.com/weather/hourbyhour/1/USNJ0524:1:US" %>%
read_html() %>% html_table(fill = TRUE)
```

 - i. Create three separate plots (3×1 layout using `facet_wrap()`) of the time of day on the x -axis against temperature, humidity, and windspeed for the day on the y -axis (using points connected by lines for each), coloring the point by the chance of precipitation. Adjust the vertical scales accordingly.
- Hint: to use `facet_wrap()` for the plots, first untidy the data using `gather()` to create another column called `variable` which specifies the temperature, humidity, and windspeed. Then use `facet_wrap(~variable, scales = "free")` to create the plots.
3. Using the code from `week6.R` on Canvas, extract the rank, views, length, and post date of each video on Youtube's trending page <https://www.youtube.com/feed/trending>. Call the rank `rank` and the number of views `views`

- (a) Transform the post date so that it is in days. For example “1 week ago” would become “7”, and “3 hours ago” would become “0.125”. Call this new variable **post_date**.
- (b) Transform the length so that it is in minutes. For example, “5:30” would become “5.5”. Call this new variable **length**.
- (c) Create a new variable **popularity** equal to **views** divided by **post_date**. Create a plot of **rank** and **popularity**.