Lecture 5

What are we doing without Mike?

- Learning how to read data from external files
  - So then we can find interesting things in the data through code
- How to write that data to a file too!
- This is going to be rough today -- but this is a really important next step
  - Visualizations/analysis are built from being able to read in data to do things with
- This will also be somewhat hands-off on our part
  - Use Google, StackOverflow, read documentation before asking questions

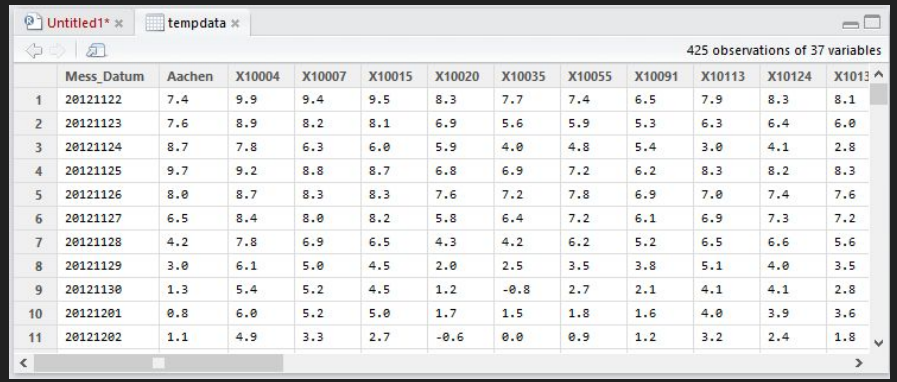
In-class exercises

You know what to do (fork and clone):

https://github.com/INFO-498F/lecture-5-exercises

Data Frames

- What is a data frame?
  - Read the README in the repo for a general idea
    - ...but it's really just a big list of vectors (and you already know how to work with vectors!)
  - You can quickly and programmatically access the columns and rows to write code to do powerful analysis
  - MORE DETAIL WILL COME FRIDAY!



Untitled1* × tempdata ×

425 observations of 37 variables

	Mess_Datum	Aachen	X10004	X10007	X10015	X10020	X10035	X10055	X10091	X10113	X10124	X1013
1	20121122	7.4	9.9	9.4	9.5	8.3	7.7	7.4	6.5	7.9	8.3	8.1
2	20121123	7.6	8.9	8.2	8.1	6.9	5.6	5.9	5.3	6.3	6.4	6.0
3	20121124	8.7	7.8	6.3	6.0	5.9	4.0	4.8	5.4	3.0	4.1	2.8
4	20121125	9.7	9.2	8.8	8.7	6.8	6.9	7.2	6.2	8.3	8.2	8.3
5	20121126	8.0	8.7	8.3	8.3	7.6	7.2	7.8	6.9	7.0	7.4	7.6
6	20121127	6.5	8.4	8.0	8.2	5.8	6.4	7.2	6.1	6.9	7.3	7.2
7	20121128	4.2	7.8	6.9	6.5	4.3	4.2	6.2	5.2	6.5	6.6	5.6
8	20121129	3.0	6.1	5.0	4.5	2.0	2.5	3.5	3.8	5.1	4.0	3.5
9	20121130	1.3	5.4	5.2	4.5	1.2	-0.8	2.7	2.1	4.1	4.1	2.8
10	20121201	0.8	6.0	5.2	5.0	1.7	1.5	1.8	1.6	4.0	3.9	3.6
11	20121202	1.1	4.9	3.3	2.7	-0.6	0.0	0.9	1.2	3.2	2.4	1.8

How do we read data in?

- If you have a file you want to read in (e.g. a **.csv** file):
  - **read.csv(“path/to/myfile.csv”)**
  - What’s a .csv?
    - File type that stands for “Comma Separated Values” - a file that has a bunch of values separated by commas (wow)
    - example.csv:
    - name,age,city
    - marcus,21,seattle
    - alex,20,new york
    - jordan,20,dallas
    - iman,20,seattle
  - R reads through every line in a CSV file and creates a data frame that you can use

How to use that data you read in

- Access a column using $

  > data$tas

  [1] “Marcus” “Alex” “Jordan” “Iman”

- Save the column as a variable for quick access to that column

  > ta_names <- data$tas

Vector vs. Factor

- Sometimes data frames will store columns as factors instead of a vector
- The difference between a vector and a factor isn't too important for this lecture (find more information in the README.md file of the repo)
- We can do a simple check to see if it is a vector or not:

```
> is.vector(data$tas)

# returns FALSE
```

- If it isn't a vector, let R know that you want a vector:

```
> ta_names <- as.vector(data$tas) # hooray, it's a vector
```

Exercise 1

1. **setwd()**: Lets R know where you are working
   a. Use absolute path: “C:/Users/Alex/Desktop/lecture-5-exercises”
2. **View()**: Lets you look at the data that you pass in
3. **is.vector()**: Returns a boolean indicating if the argument is a vector
4. **Find specific element in dataframe:**

```
> age <- data$age
> twenty_one_years <- ta_names[age == 21]
```

Exercise 2

Exercise 3

Help, I need to to know exactly what functions to use on this exercise!

Use Google, or **help()** on the following functions:

- **colnames()**
- **write.csv()**
- **as.character()**