

R PROGRAMMING

Intermediate Skills






INTRODUCTIONS

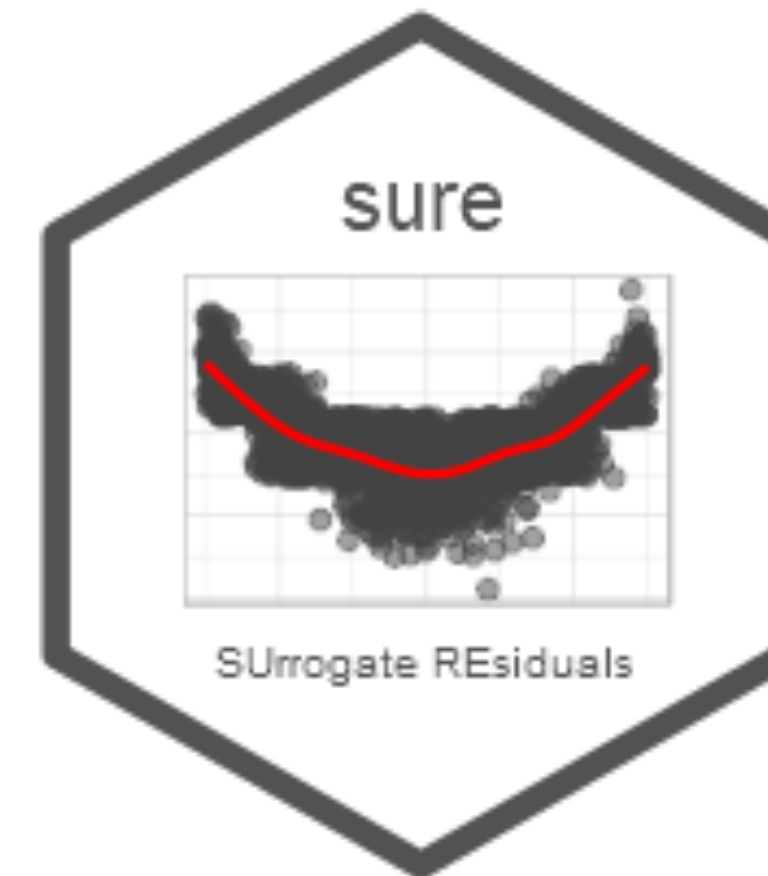
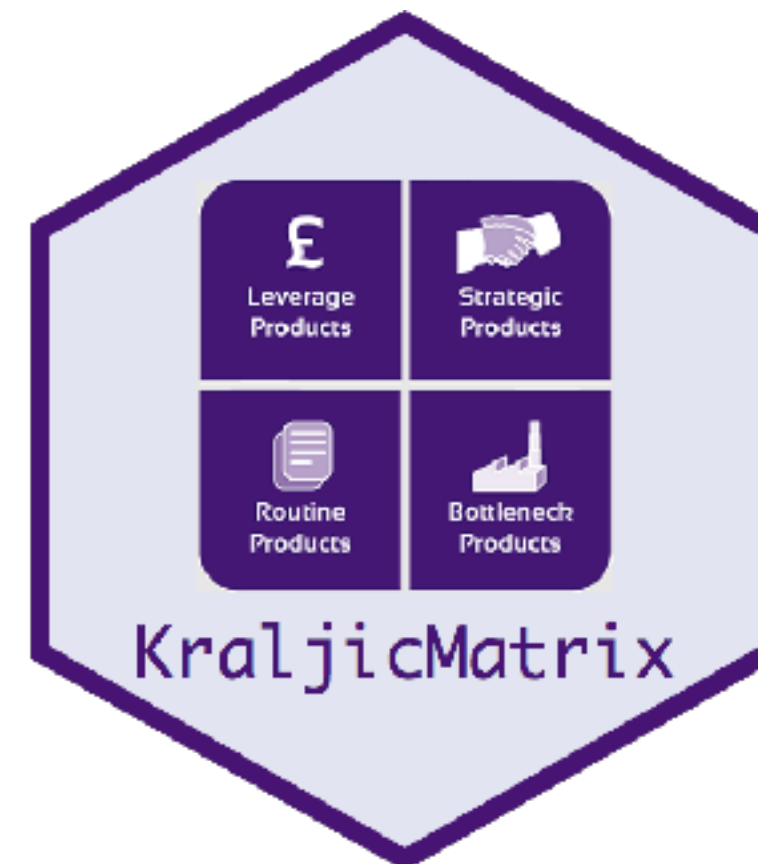
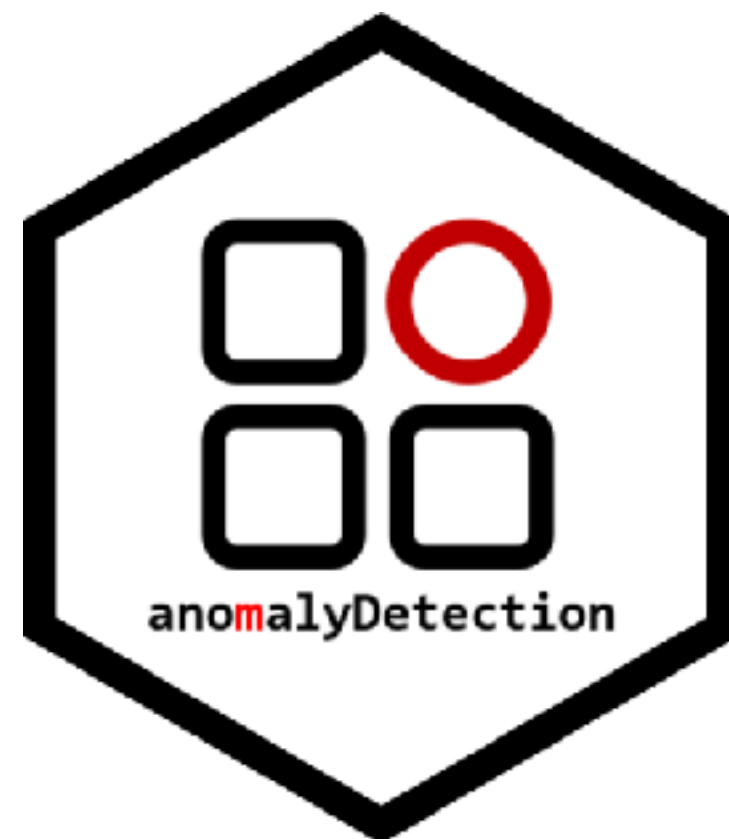
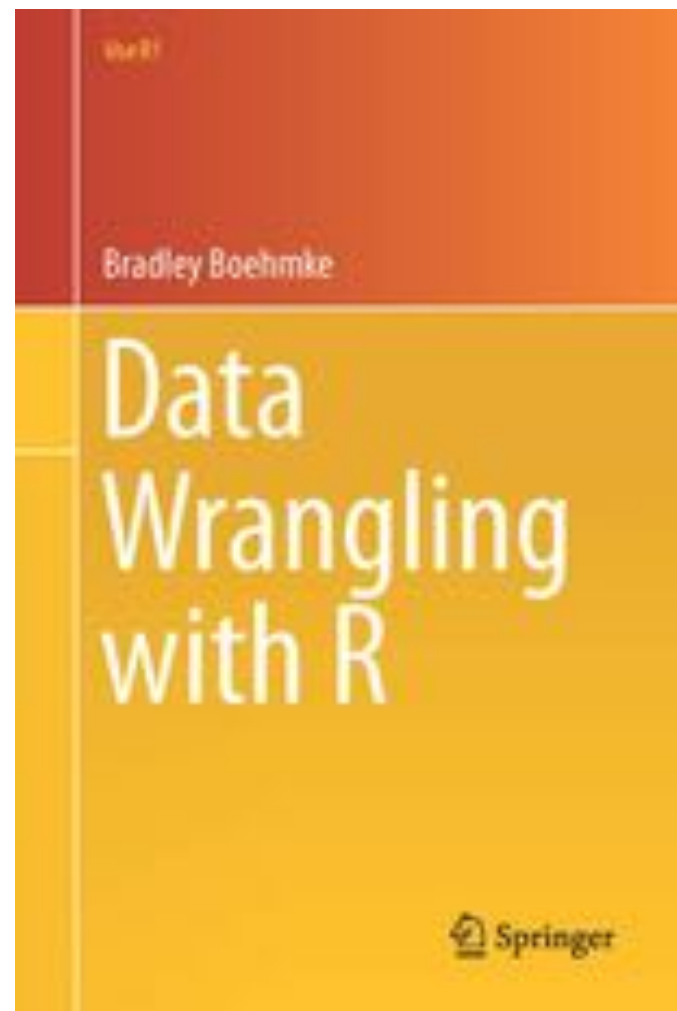
Who is this guy?

A faded background image of a family at a park. On the left, a man in a blue shirt and jeans sits on a red and purple exercise bike. On the right, a man in a blue shirt and sunglasses stands with a woman in a black t-shirt and sunglasses, and two young girls. They are all smiling and appear to be enjoying a day outdoors.

HELLO
my name is

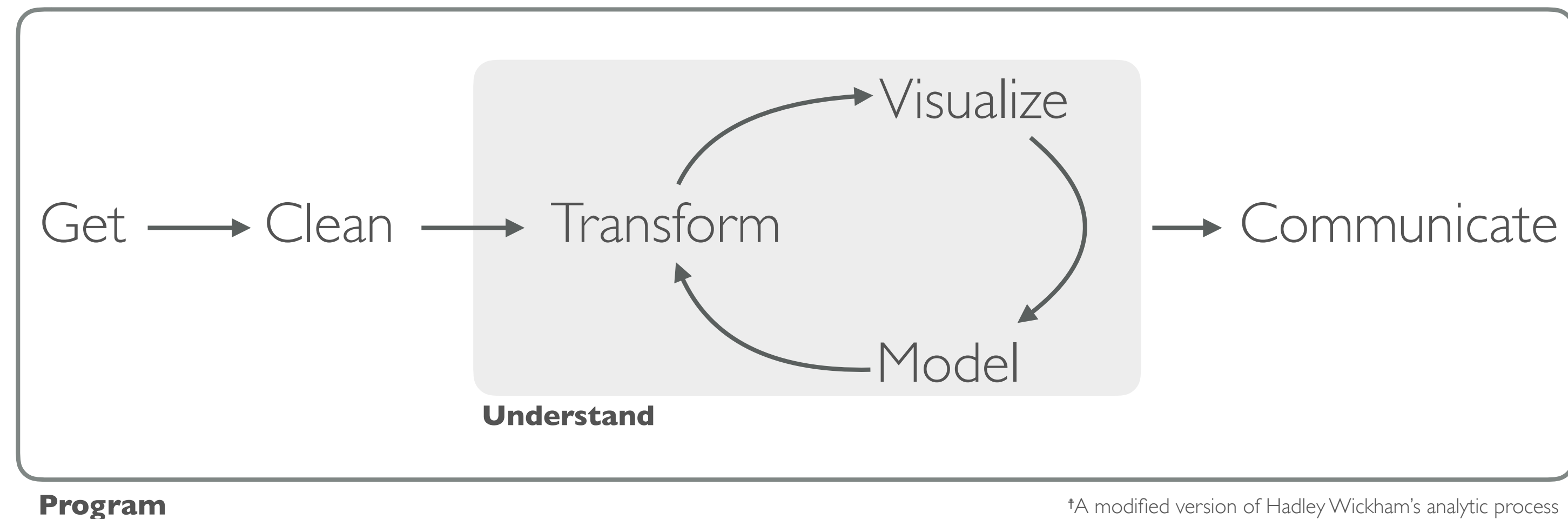
Brad

-  bradleyboehmke.github.io
-  bradleyboehmke@gmail.com
-  [@bradleyboehmke](https://twitter.com/bradleyboehmke)
-  [bradleyboehmke](https://github.com/bradleyboehmke)
-  [bradleyboehmke](https://www.linkedin.com/in/bradleyboehmke)



SETTING THE EXPECTATIONS

- Introduction to R
- **Intermediate R** 🧑
- Text Mining with R (March 8-9)
- Applied Analytics with R
- Machine Learning with R (May 14-15)



†A modified version of Hadley Wickham's analytic process

SETTING THE EXPECTATIONS

Day 1

- Reproducibility & reporting
- Data transformation & mutation
- Joining relational data

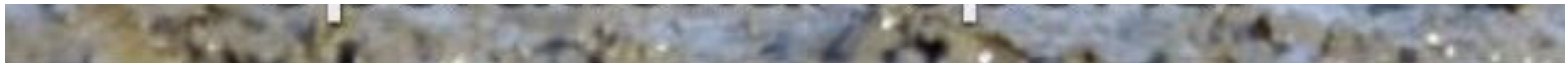
Day 2

- Control statements & iteration
- Writing functions
- Intro to modeling

Efficiency→ *Analytics*

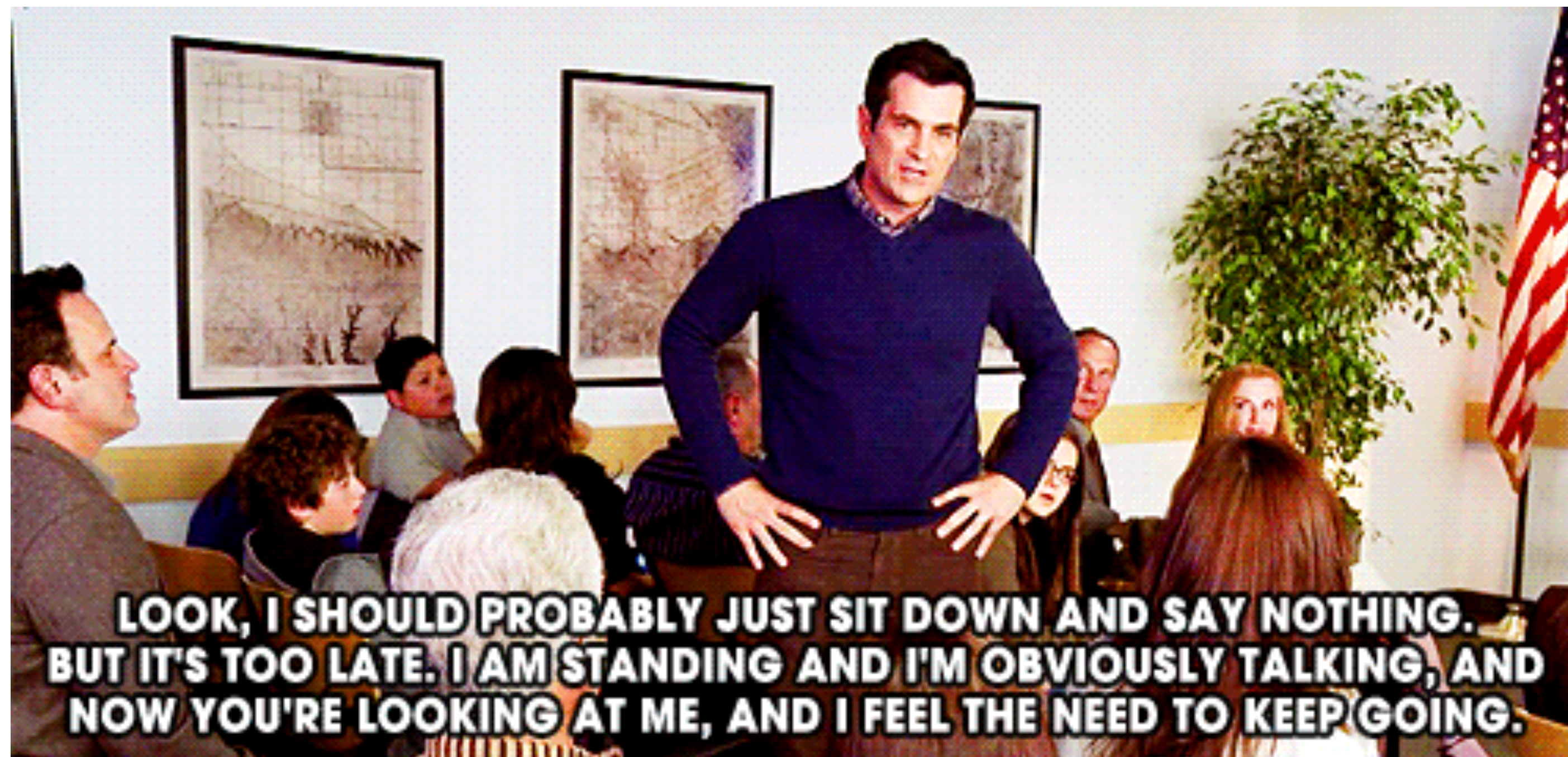


You will be overwhelmed!



My Teaching Philosophy

THIS IS MEANT TO BE A DISCUSSION



YOUR TURN!

Lots of hands-on coding exercises

Strong proponent of
collaborative work!



YOUR TURN!

Introduce yourself to your neighbors:

Who are you and what is your experience with R?

Regarding the topics that will be covered what are your strengths? Weaknesses?

WARM-UPS

Enough chit-chatting, time to code!



VECTOR EXERCISES

1. check out the built-in vector **state.region**
2. what type of data does this contain?
3. Subset for only north central states. How many north central states are there?
4. Change `state.name` to a character variable, add **state.region** values as names to the **state.name** vector, subset for north central states.

MATRIX EXERCISES

1. check out the built-in **VADeaths** matrix data

2. what attributes does **VADeaths** have?

3. Calculate averages for each column and row

4. Can you figure out how to add these averages to your table so the output looks like:

| | Rural Male | Rural Female | Urban Male | Urban Female | Avg_by_Age |
|--------------|------------|--------------|------------|--------------|------------|
| 50-54 | 11.70 | 8.70 | 15.40 | 8.40 | 11.050 |
| 55-59 | 18.10 | 11.70 | 24.30 | 13.60 | 16.925 |
| 60-64 | 26.90 | 20.30 | 37.00 | 19.30 | 25.875 |
| 65-69 | 41.00 | 30.90 | 54.60 | 35.10 | 40.400 |
| 70-74 | 66.00 | 54.30 | 71.10 | 50.00 | 60.350 |
| Avg_by_Local | 32.74 | 25.18 | 40.48 | 25.28 | 30.920 |

DATA FRAME EXERCISES

1. Load the `nycflights13` package
2. Using the `flights` data, select the first 1000 rows and the following columns: `month`, `dep_delay`, `carrier`, `distance`, `time_hour`. Save this as `small_flights`
3. Look at the structure and summary of `small_flights`
4. Rename the columns of `small_flights` to "Month", "Delay", "Carrier", "Distance", "Date-Time"
5. Look at the first and last 15 rows

LIST EXERCISES

1. *Create this regression model:*

```
flight_lm <- lm(arr_delay ~ dep_delay + month + carrier,  
               data = flights)
```

2. *What list items does **flight_lm** contain?*

3. *Extract the residuals from the **flight_lm** list*

4. *What is the min, max, median, and mean of these residuals?*

LET'S GET STARTED!

