Applied Statistical Analysis

EDUC 6050

Week 1

Finding clarity using data

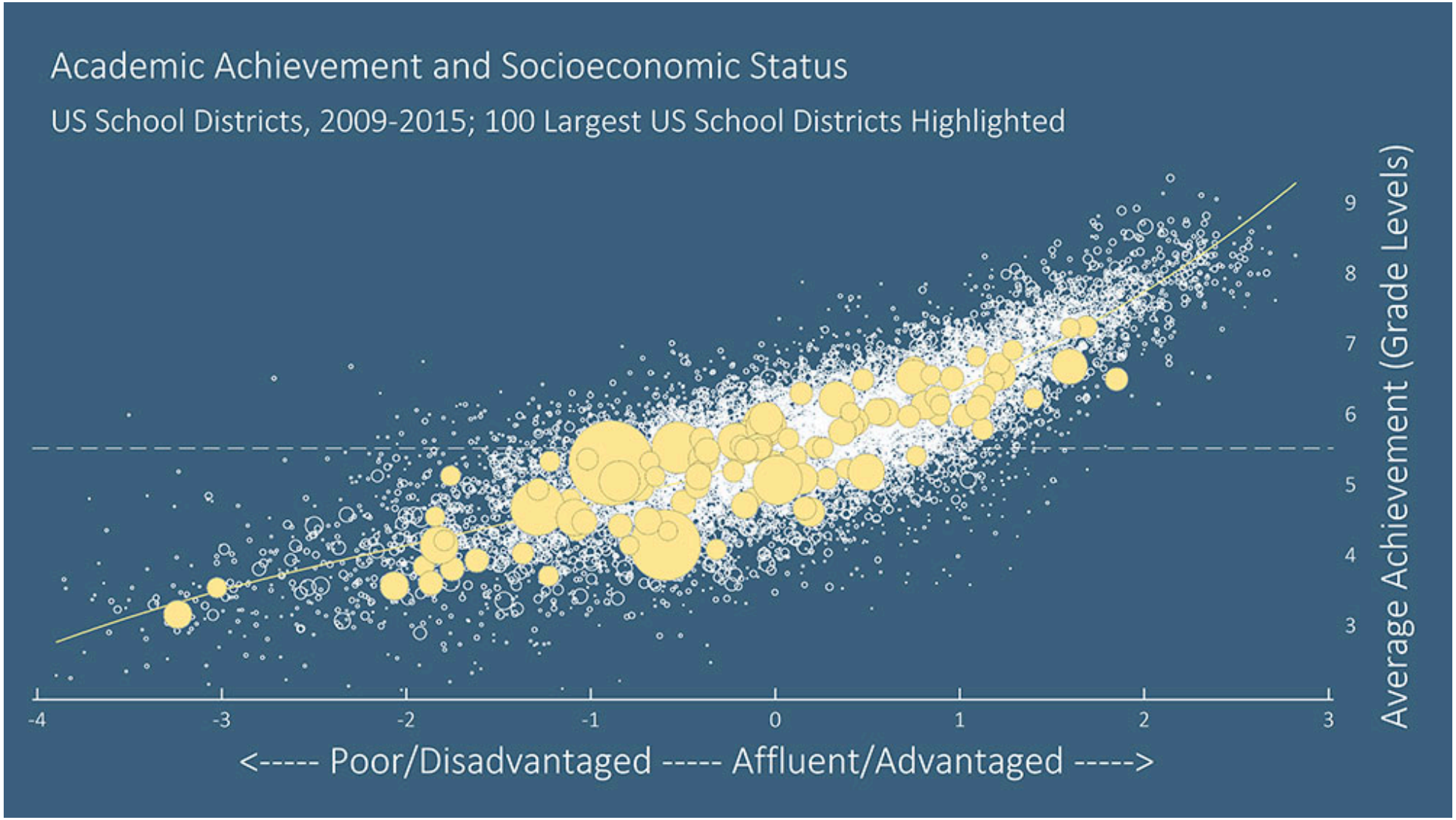
Welcome

1. What is quantitative research?
2. How does data inform our world?
3. How are data analyzed?

Data, Data, Data, Data, Data, ...

Tesla Autopilot

Data, Data, Data, Data, Data, ...



Data, Data, Data, Data, Data, ...

Health Care Policy and Cost

Data are/is Cool

“In God we trust. All others must bring data.”
W. Edwards Deming

“It is a capital mistake to theorize before one has data.”
Sherlock Holmes, “A Study in Scarlett” (Arthur Conan Doyle).

“You can have data without information, but you cannot have information without data.”
Daniel Keys Moran

Purpose of this course

Develop quantitative understanding and skills

Prepare you for:
1. Your thesis
2. Your career

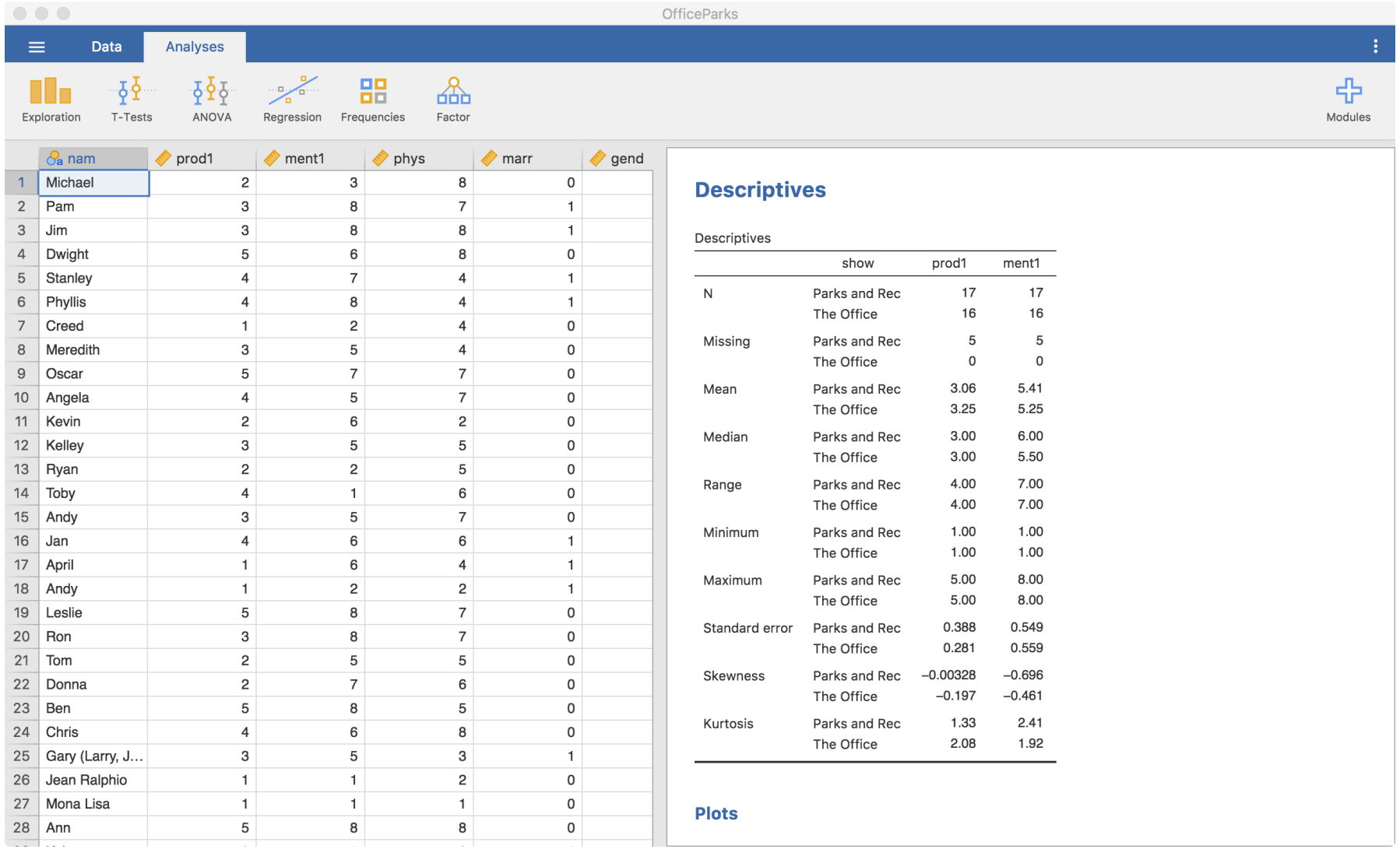
What is expected of you

- Attend and participate in class
- Prepare for class (readings before class)
- Professional correspondence with colleagues
- Use assignments to learn
- Ask questions
- Communicate with me

Syllabus

https://tysonbarrett.com/syllabus/asa

Jamovi



	nam	prod1	ment1	phys	marr	gend
1	Michael	2	3	8	0	
2	Pam	3	8	7	1	
3	Jim	3	8	8	1	
4	Dwight	5	6	8	0	
5	Stanley	4	7	4	1	
6	Phyllis	4	8	4	1	
7	Creed	1	2	4	0	
8	Meredith	3	5	4	0	
9	Oscar	5	7	7	0	
10	Angela	4	5	7	0	
11	Kevin	2	6	2	0	
12	Kelley	3	5	5	0	
13	Ryan	2	2	5	0	
14	Toby	4	1	6	0	
15	Andy	3	5	7	0	
16	Jan	4	6	6	1	
17	April	1	6	4	1	
18	Andy	1	2	2	1	
19	Leslie	5	8	7	0	
20	Ron	3	8	7	0	
21	Tom	2	5	5	0	
22	Donna	2	7	6	0	
23	Ben	5	8	5	0	
24	Chris	4	6	8	0	
25	Gary (Larry, J…	3	5	3	1	
26	Jean Ralphio	1	1	2	0	
27	Mona Lisa	1	1	1	0	
28	Ann	5	8	8	0	

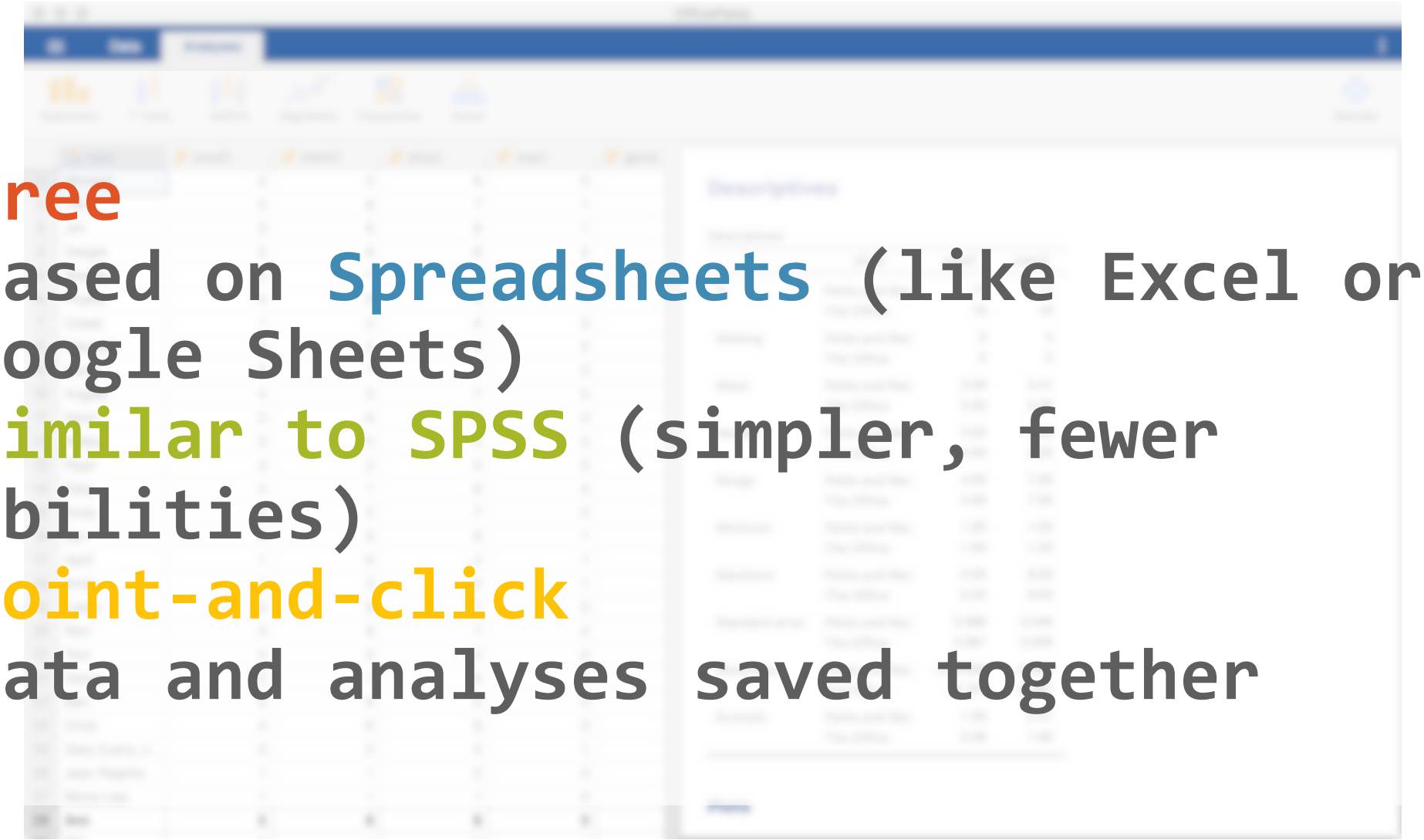
Descriptives

Descriptives

	show	prod1	ment1
N	Parks and Rec	17	17
	The Office	16	16
Missing	Parks and Rec	5	5
	The Office	0	0
Mean	Parks and Rec	3.06	5.41
	The Office	3.25	5.25
Median	Parks and Rec	3.00	6.00
	The Office	3.00	5.50
Range	Parks and Rec	4.00	7.00
	The Office	4.00	7.00
Minimum	Parks and Rec	1.00	1.00
	The Office	1.00	1.00
Maximum	Parks and Rec	5.00	8.00
	The Office	5.00	8.00
Standard error	Parks and Rec	0.388	0.549
	The Office	0.281	0.559
Skewness	Parks and Rec	−0.00328	−0.696
	The Office	−0.197	−0.461
Kurtosis	Parks and Rec	1.33	2.41
	The Office	2.08	1.92

Plots

Jamovi



- Free
- Based on Spreadsheets (like Excel or Google Sheets)
- Similar to SPSS (simpler, fewer abilities)
- Point-and-click
- Data and analyses saved together

Tell us about yourself

https://docs.google.com/spreadsheets/d/1JaggNgUtkdzQ9T-FTNDsvf0D6DRM1psMJZvcKuYSaxI/edit?usp=sharing

- first_name: your first name
- degree: the degree you're pursuing
- grow_up: what you want to do when you grow up
- hobby: one of your hobbies
- where_from: where are you from?
- would_rather: fly or money?
- amount_netflix: how much time per week spent watching Netflix
- rating_office: rating 0-10 of The Office (US)

Tell us about yourself

https://docs.google.com/spreadsheets/d/1JaggNgUtkdzQ9T-FTNDsvf0D6DRM1psMJZvcKuYSaxI/edit?usp=sharing

- first_name: your first name
- course: the course you are pursuing
- grow_up: what you want to do when you grow up
- hobby: one of your hobbies
- where_from: where are you from?

We will practice with Jamovi next week using this data

- would_rather: fly or money?
- amount_netflix: how much time per week spent watching Netflix
- rating_office: rating 0-10 of The Office (US)

Data and spreadsheets

Get used to working with spreadsheets

- Excel, Google Sheets, Numbers, etc.

Good data practices

- Save a **master data file** that does not change after you have cleaned up the file
- Do NOT save subsetted data files (e.g., removed all ages < 20)
  - Instead **save the analyses**
- Save the master file on **multiple devices** (flash drive, cloud, computer)

Broman et al. (2017)

Good data practices

- **Double check** your work
  - Re-run the same analyses after closing down the file and software
- **Keep track** of all your data and analysis

Broman et al. (2017)

Assignments require your own data

You can use any data that you'd like, if:

- It has both **continuous** and **categorical** variables
- It has at least **20 participants**
- It is available for you to use in class (**de-identified**)

Assignments require your own data

Continuous is where the values of the variable can be a wide, continuous range

- It has both **continuous** and **categorical** variables
- It has at least **20 participants**
- It is available for you to use in class (de-identified)

Assignments require your own data

You can use any data that you'd like, if:

Categorical is where the values of the variable can only be a few, predefined values

- It has both continuous and categorical variables
- It has at least 20 participants
- It is available for you to use in class (de-identified)

Questions?

Next week:

1. Working with Data
2. Overview of Statistics
3. Intro to Statistical Terminology
4. Intro to Jamovi