

Applied Statistical Analysis

EDUC 6050

Week 1

Finding clarity using data

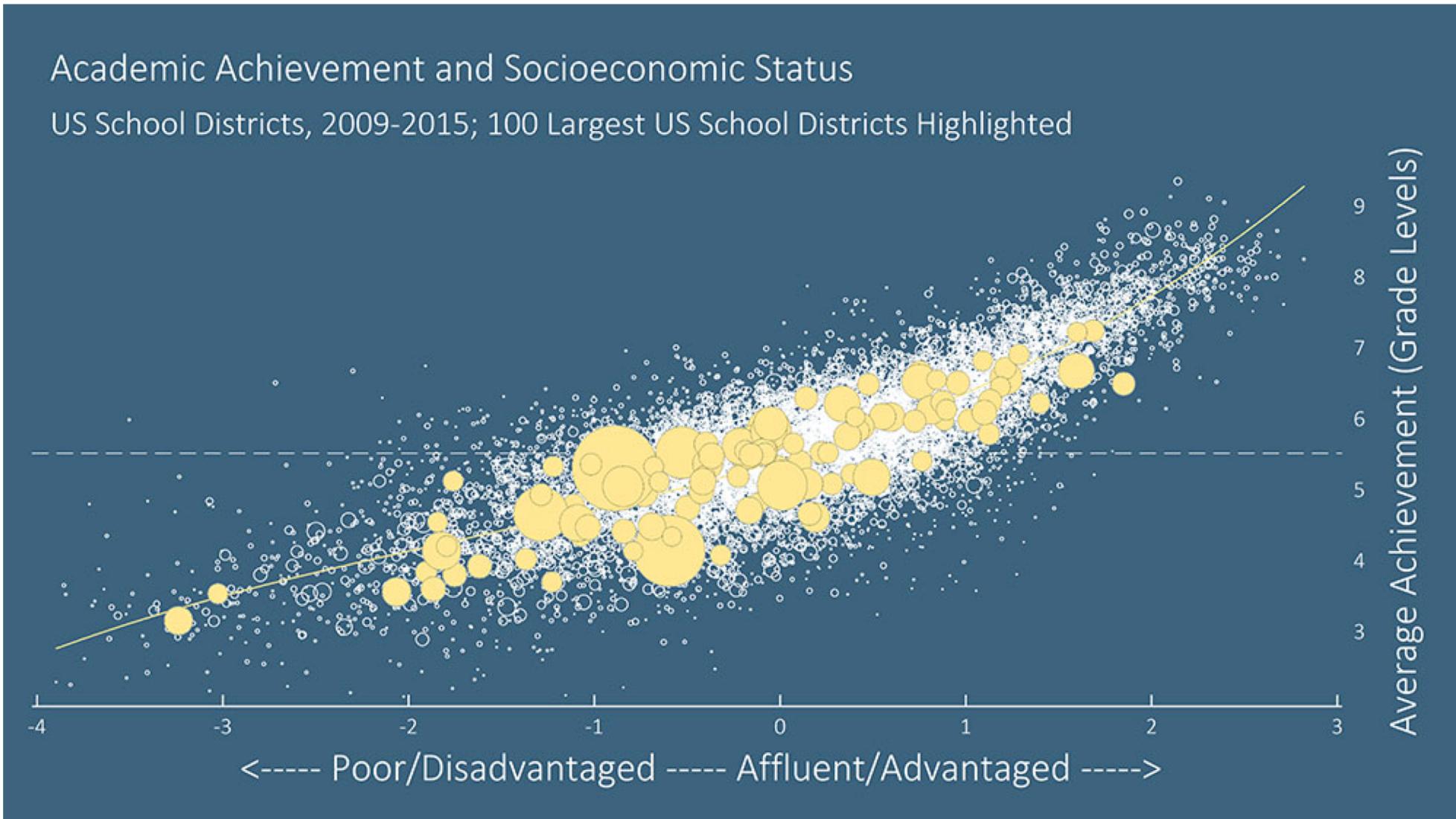
Welcome

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 Scarlet” (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

<http://tysonbarrett.com/syllabus/asa>

Jamovi

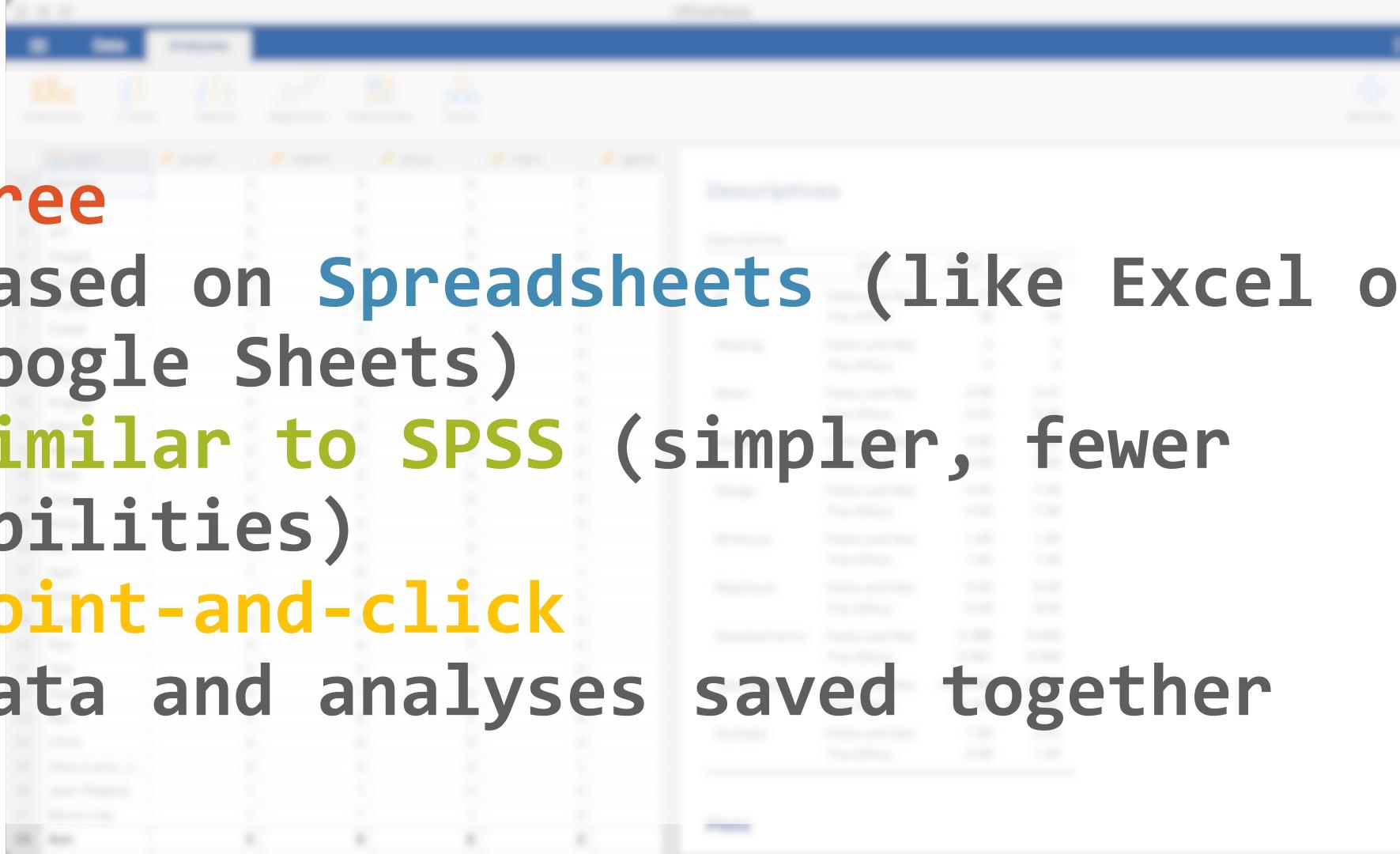
The screenshot shows the Jamovi interface with the following details:

- Top Bar:** Shows the title "OfficeParks" and tabs for "Data" (selected) and "Analyses".
- Analyses Tab:** Displays icons for Exploration, T-Tests, ANOVA, Regression, Frequencies, and Factor.
- Data View:** A table with columns: nam, prod1, ment1, phys, marr, and gend. The first row (Michael) is selected. The data rows range from 1 to 28, listing names and scores for each variable.
- Descriptives Analysis:** A table titled "Descriptives" comparing two groups: "Parks and Rec" and "The Office".
- Descriptives Table:**

	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:** A section labeled "Plots" is visible at the bottom.

Jamovi

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



Schedule

- Tentative
- Readings are due before class
- Assignment by the end of the day

Date	Readings and Such	Lecture Topic	Week	Assignment Due by 11:59pm
Jan 9		Syllabus, Textbook, data, and create your survey	1	
Jan 16	Broman et al. (2017) – only sections 2-4,6-8	Working with and Analyzing Data, Overview of Statistics, Intro to Statistics Terminology, Introduction to Jamovi	2	
Jan 23	Ch 2, 3 Start looking for published research in your area	Statistics terminology (Hypothesis, IV and DV, Measurement, Validity and Reliability, Correlation and Experimentation, Distributions, Central Tendency and Variability)	3	
Jan 30	Ch 4, 5, 6	Statistics terminology continued (hypothesis testing, populations and samples, descriptive and inferential statistics, effect sizes, confidence intervals, Type I and II errors)	4	
Feb 6	Ch 7	More on Jamovi (data manipulation, transformations, assumptions), Creating tables and figures for reports and manuscripts, Intro to t-tests	5	HW #1 (Central Tendency and Variability)
Feb 13	Ch 7, 9, 10	T-tests (student's, Mann-Whitney, Wilcoxon), Review of hypothesis tests	6	
Feb 20	Ch 11, 12	ANOVA (one-way, two-way), ANCOVA, Repeated Measures ANOVA, post-hoc analyses	7	HW #2 (t-tests)
Feb 27		Mid-Term Examination	8	HW #3 (ANOVA)
Mar 6	Ch 13	Correlations (Pearson, Spearman, partial)	9	
Mar 13		Spring Break! (Do Not Come to Class)		
Mar 21	Ch 13	Linear Regression (hypothesis testing, prediction, assumptions)	10	
Mar 27	Ch 13	Multiple Regression (moderation, mediation)	11	HW #4 (correlations, regression)
April 3	Ch 14	Categorical Data Analysis (Chi-square, logistic, log-linear, odds ratios)	12	HW # 5 (multiple regression)
April 10	Ch 14	Categorical Data Analysis continued (logistic, odds ratios)	13	
April 17		Research Portfolio, Review for final	14	HW #6 (categorical data)
April 24		Review (get ready for the final)	15	
May 1		Final Examination		

Schedule

- Tentative
- Readings are due before class
- Assignments are due by the end of the day

Please read the syllabus in depth

Date	Readings and Such	Lecture Topic	Week	Assignment Due by 11:59pm
Jan 9		Syllabus, Textbook, data, and create your survey	1	
Jan 16	Broman et al. (2017) – only sections 2-4,6-8	Working with and Analyzing Data, Overview of Statistics, Intro to Statistics Terminology, Introduction to Jamovi	2	
Jan 23	Ch 2, 3 Start looking for published research in your area	Statistics terminology (Hypothesis, IV and DV, Measurement, Validity and Reliability, Correlation and Experimentation, Distributions, Central Tendency and Variability)	3	
Jan 30	Ch 4, 5, 6	Statistics terminology continued (hypothesis testing, populations and samples, descriptive and inferential statistics, effect sizes, confidence intervals, Type I and II errors)	4	
Feb 6	Ch 7	More on Jamovi (data manipulation, transformations, assumptions), Creating tables and figures for reports and manuscripts, Intro to t-tests	5	HW #1 (Central Tendency and Variability)
Feb 13	Ch 7, 9, 10	T-tests (student's, Mann-Whitney, Wilcoxon), Review of hypothesis tests	6	
Feb 20	Ch 11, 12	ANOVA (one-way, two-way), ANCOVA, Repeated Measures ANOVA, post-hoc analyses	7	HW #2 (t-tests)
Feb 27	Midterm	Midterm	8	HW #3 (ANOVA)
March 5	Ch 13	Correlations (Pearson, Spearman, Gamma)	9	
Mar 13		Spring Break! (Do Not Come to Class)		
Mar 21	Ch 14	Linear Regression (hypothesis testing, predictions, assumptions)	10	
Mar 27	Ch 15	Multiple Regression (moderation, mediation)	11	HW #4 (correlations, regression)
April 3	Ch 14	Categorical Data Analysis (Chi-square, logistic, log-linear, odds ratios)	12	HW # 5 (multiple regression)
April 10	Ch 14	Categorical Data Analysis continued (logistic, odds ratios)	13	
April 17		Research Portfolio, Review for final	14	HW #6 (categorical data)
April 24		Review (get ready for the final)	15	
May 1		Final Examination		

Break Time

Go to next slide

Tell us about yourself

During the break, please go to:

<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?
- **where_end**: where do you want to end up living
- **would_rather**: fly or money?

Tell us about yourself

During the break, please go to:

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

We will practice with Jamovi next

- **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?
- **where_end:** where do you want to end up living
- **would_rather:** fly or money?

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)

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

Break Time

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 it's categorical and has at least 20 participants.

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)

Create a survey

Use a **survey** to collect data that you can use for this class

- Needs to meet the requirements of the data

Example

Questions?

Next week:

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