

Group advice

Working on group projects allows you to build teamwork skills, learn how to distribute workloads and responsibilities, and navigate different points of view. These are all valuable skills you will use in academics and your future career.

You can take steps to have a good group experience from the start. Here are tips for a positive and productive group project.

Assign responsibilities

At the beginning of the project, set up a meeting to get to know your group members. Ask each group member to review the assignment before the meeting to learn about the project and the different requirements, roles and responsibilities.

During the meeting, you can work together to decide who will be responsible for each part of the project. Consider what works best for the group based on each person's skill set.

Set deadlines and due dates

Creating a project timeline can help your group keep things on track and avoid misunderstandings about when each person is responsible for completing their tasks. Review the syllabus and project guidelines to determine due dates, and decide when you will have the project completed and ready to turn in or present. Then, work together to set due dates for each step of the project.

Build in time to help each other and combine work. Give yourselves time to rework any parts of the project that need extra effort or edits.

Allow time for the group to form

It can take time to settle into a productive team, and sometimes there may be conflict. However, working through disagreements can lead to a strong and healthy group. When you are assigning responsibilities and creating a project timeline, people can have different ideas about how to do things. It is OK to voice your concern. Listen to others as they share their thoughts and ideas, and work together to compromise.

Communicate effectively

Clear communication with positive intent can help groups to function productively. Communication may feel awkward if there is a disagreement or misunderstanding. Emotions can also be intense during difficult discussions, causing minor disagreements to escalate.

Communicate directly with your group members and focus on the problem instead of getting caught up in the emotions. Also, address any conflicts or difficult conversations with the people involved and avoid complaining about people who are not present. Remember to use “**I**” statements to describe how the situation makes you feel and try framing the issue neutrally to decrease any feelings of blame. Here are some examples:

- “I feel frustrated when parts of the project are incomplete because we all have a responsibility to help.”
- “I feel hurt when my ideas are not taken seriously because I think they could improve our report.”
- “I am upset that a decision was made without group consensus because we want this to be a collaborative project.”

When sharing your point of view, speak from your perspective and avoid making assumptions. Avoid bringing up past issues or making personal attacks.

Also, be sure that the conversation is a two-way discussion. Give your classmate space to share their point of view by acknowledging how they feel about the situation. Listen fully to what they are saying without judgment and ask questions to make sure you understand what they are trying to convey. This approach can help them feel heard and help you both find a solution.

Form connections with others

Strengthening your relationships with your group members can help with productivity. It allows you to focus on the task at hand and work together to find a solution to any challenges. It may also enable you to develop empathy and interact with more kindness toward each other.

If possible, meet in person versus communicating only online. For example, you can sit together in class or meet up for a meal in the dining hall to discuss updates on your project. Efforts to have a personal connection can make working on your project more enjoyable and expand your network.

Celebrate your accomplishments

Once you’ve completed the group project, remember to celebrate your group’s success. Treat yourself and recognize the hard work of your peers—you deserve it!

Kinds of analysis

Jason G. Fleischer, Ph.D.

Asst. Teaching Professor

Department of Cognitive Science, UC San Diego

jfleischer@ucsd.edu

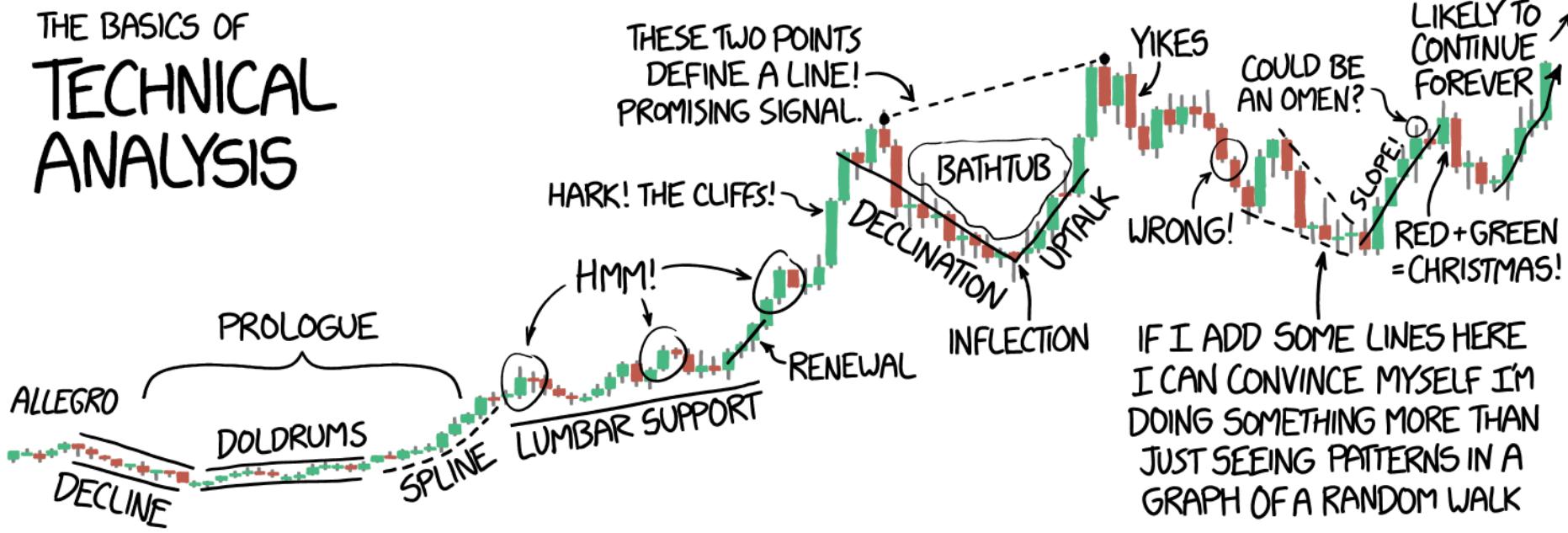


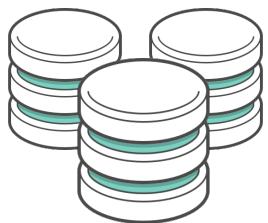
@jasongfleischer

<https://jgfleischer.com>

Slides in this presentation stolen shamelessly
from Kyle Shannon and Shannon Ellis

THE BASICS OF TECHNICAL ANALYSIS





Data

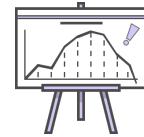


let me show you

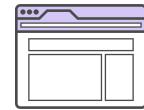
How?



A Modell!



Results!



Product!



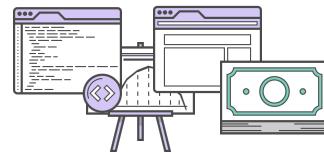
Revenue!



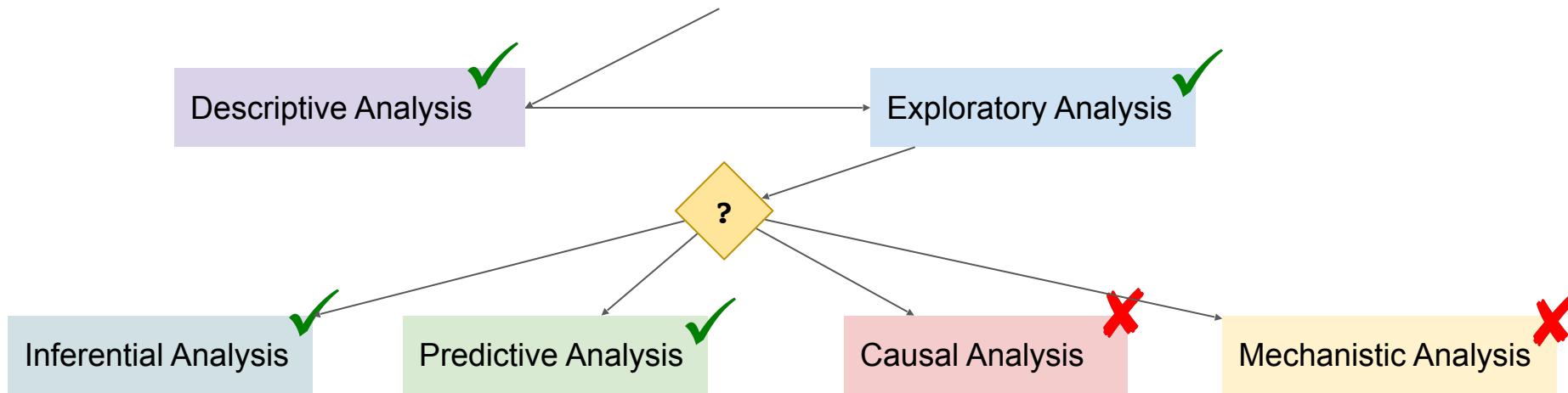
Data



The Analytic Approach Your Tool Box



The Goods

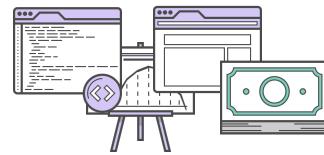




Data



The Analytic Approach Your Tool Box



The Goods

Descriptive Analysis ✓

Exploratory Analysis ✓

?

Inferential Analysis ✓

Predictive Analysis ✓

Causal Analysis ✗

Mechanistic Analysis ✗

Classic Statistics (parametric & nonparametric)

Frequentist & Bayesian

Text & Geospatial Analysis

Statistical learning/ML
- Supervised
- Unsupervised

Monte Carlo simulations

variable X \uparrow
causes

variable Y \downarrow

e.g. effects of new medication on some illness by randomized trial

variable X \uparrow 3.2 units
results in

variable Y \downarrow 1.1 units

e.g. electric current governed by wire size

Summary: Analytical Approaches

Typically Less Effort

Descriptive Analysis

- 1st thing you do on new data
- Summarize the data
- univariate plots of variables

Exploratory Analysis

- Exploring relationships
- Asking/defining questions
- univariate/bivariate/multivariate analysis and plotting
- formulate hypothesis

Inferential Analysis

- Estimating uncertainty
- test theories (infer) about the population (data gen. process)
- Building inference models

Typically More Effort

Predictive Analysis

- Building predictive models
- Use historical knowledge to predict future events
- Finding patterns

Mechanistic Analysis

- Understand precise changes one variable has on another
- typically modeled using deterministic equations
- break down complex systems into constituent parts

Causal Analysis

- Determine the average change in one variable when you alter another
- typically requires experiments (e.g. randomized studies)
- manipulate one variable observe effect on other

Exploring Analyses

General question: What impacts politics in America?

Data Science question: Is there a relationship between the sentiment of political words in South Park and America's presidential approval rating?

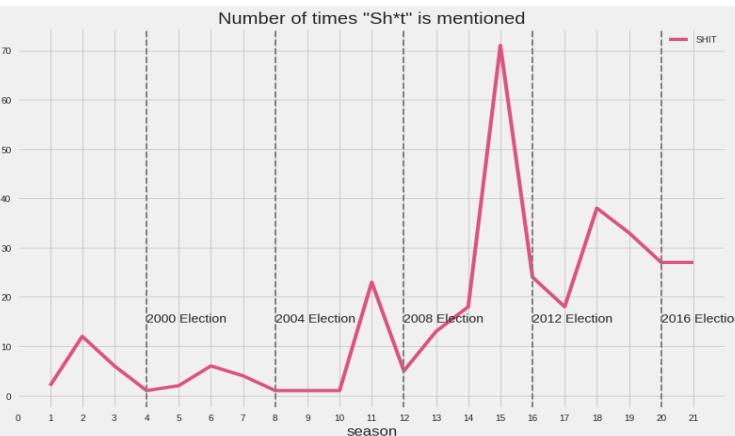
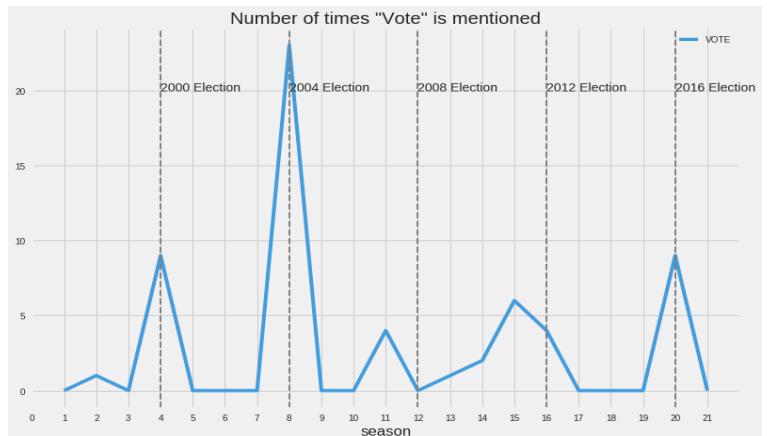
Descriptive

Exploratory

Inferential

Text Analysis

Classic Statistics
(parametric & nonparametric)



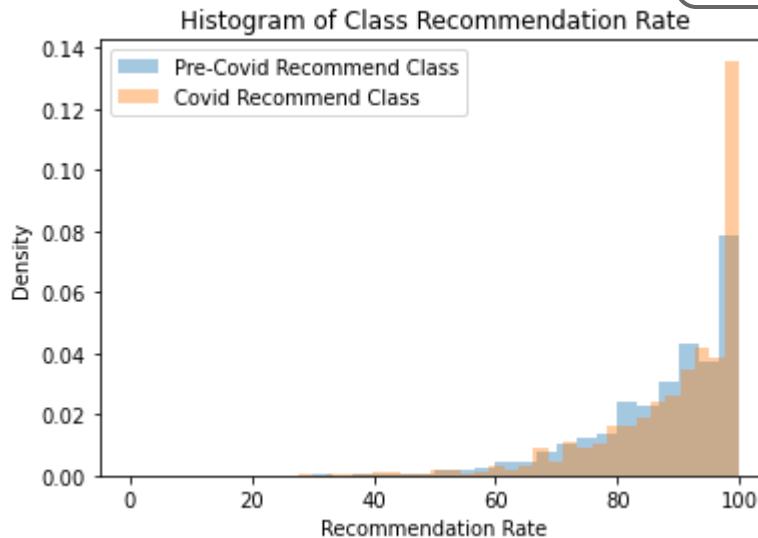
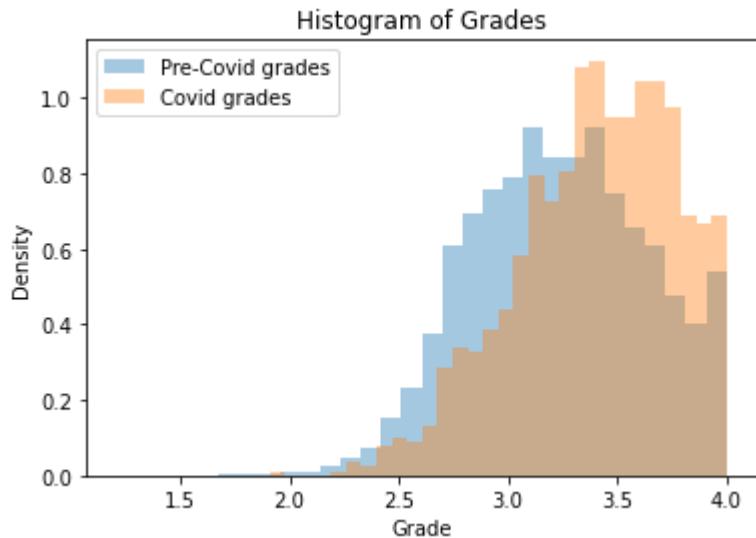
Descriptive

Exploratory

Inferential

General question: How has COVID-19 impacted students?

Data Science question: At UCSD, is there a difference between students' grades and how they rate their classes before COVID-19 and during remote learning, due to COVID-19?



Classic Statistics
(parametric &
nonparametric)

General question: Why isn't police response time always the same?

Data Science question: Where should police cars be stationed, accounting for crime levels and time of day, to make police response times equitable throughout San Diego?

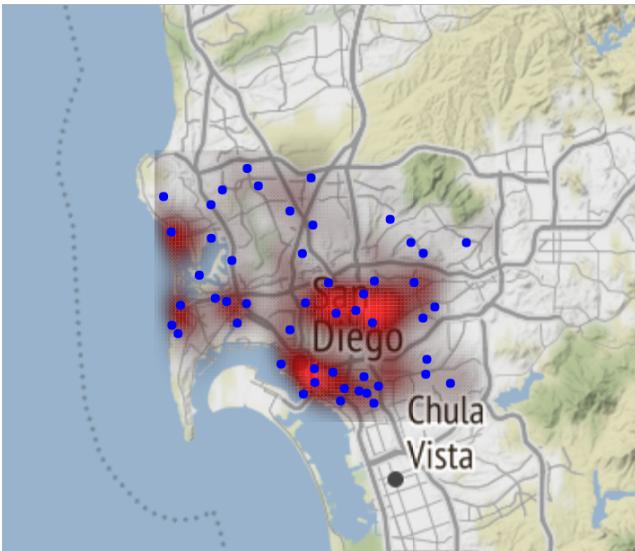
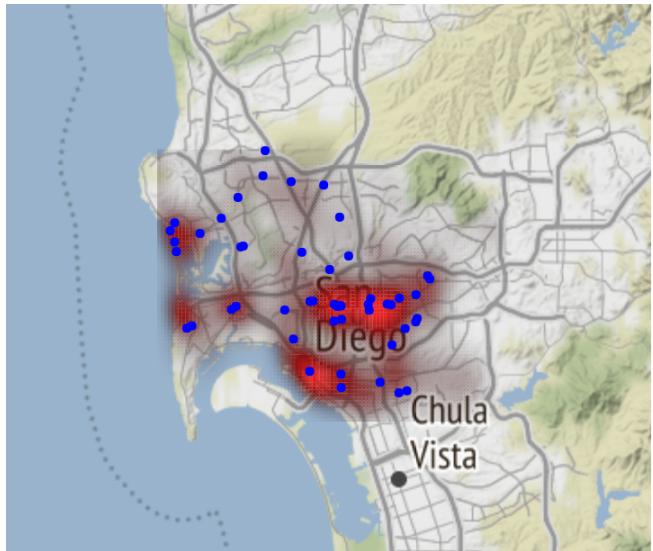
Descriptive

Exploratory

Predictive

Inferential

Geospatial Analysis



General question: What gets too much attention in the news?

Descriptive

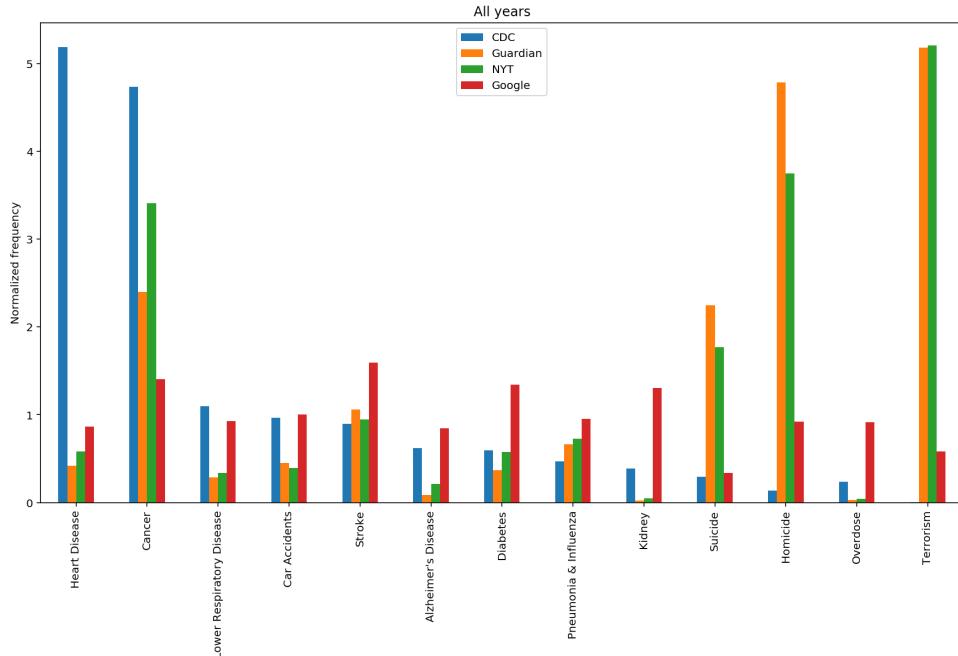
Data Science Question: Is there a relationship over time between cause of death terms in the *NYT*, The Guardian, and Google trends data relative to data from the CDC?

Exploratory

Inferential

Text Analysis

Classic Statistics
(parametric & nonparametric)

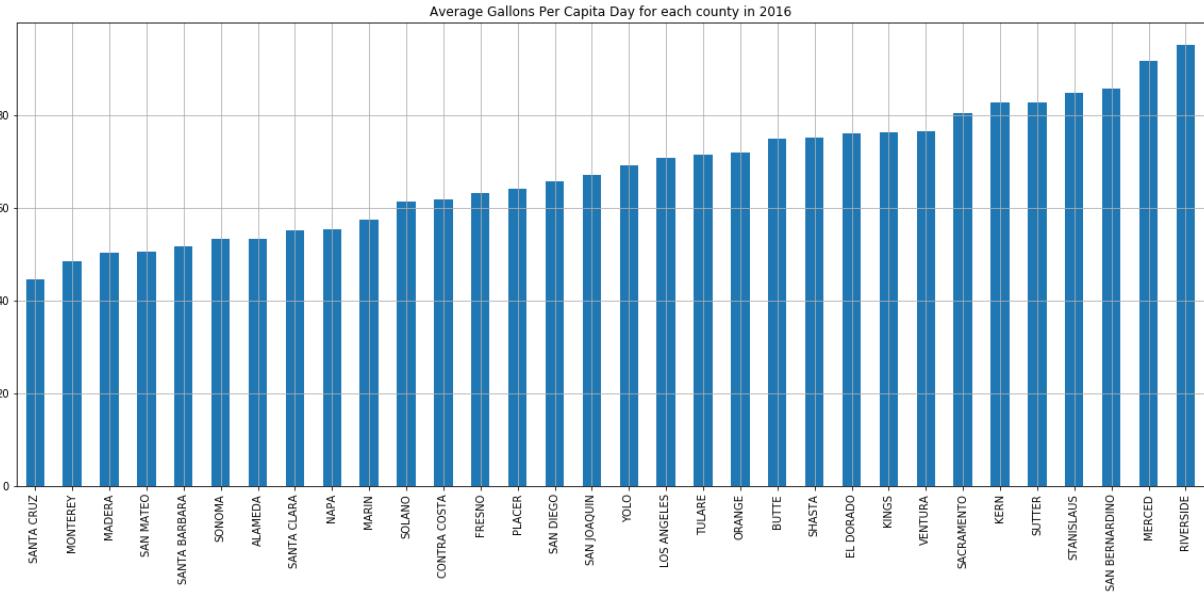


In case of the total drought in California, how many desalination plant projects we need to supply residential use water for population who live in urban areas in California?

Descriptive

Exploratory

Predictive



Descriptive and Exploratory Analysis

Jason G. Fleischer, Ph.D.

Asst. Teaching Professor

Department of Cognitive Science, UC San Diego

jfleischer@ucsd.edu

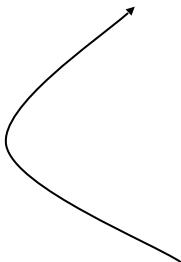


@jasongfleischer

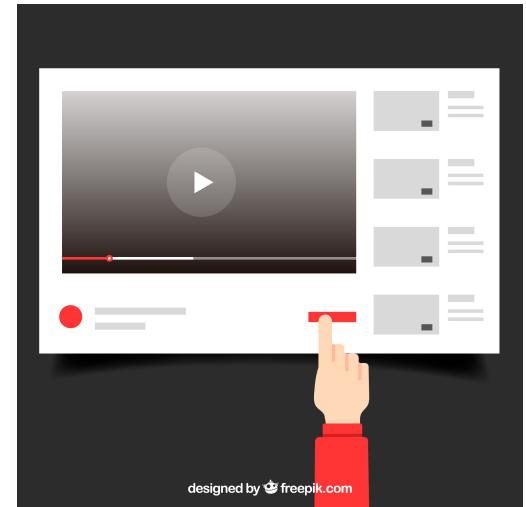
<https://jgfleischer.com>

Descriptive: The goal of descriptive analysis is to understand the components of a data set, describe what they are, and explain that description to others who might want to understand the data.

- Problem: Understanding whether users are nice or mean on YouTube
- Data science question: Are the words that people use in their comments more frequently positive words (great, awesome, nice, useful) or negative words (bad, stupid, lame, awful)?
- Type of analysis: Descriptive analysis



To answer this you would calculate statistics about YouTube comments



designed by  freepik.com

Statistics

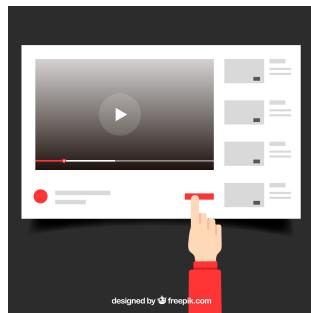
*“the science that deals with the **collection, classification, analysis, and interpretation of numerical facts or data**”*

A statistic:

“A quantity computed from a sample”

statistic

“A quantity computed from a sample”



For our YouTube analysis, we could take a random sample of comments from YouTube and calculate the following statistic: *the number of positive and the number of negative words in each review.*



Population

All comments on YouTube

During the second quarter of 2020, almost 2.13 billion comments on YouTube videos were removed due to violation of the platform's community guidelines. - J Clement on stata.com

We want to learn something about this...

Sampling

Inference

....but we can only *actually* collect data from this



1 million
comments from 2020

Sample

Best sampling practices:

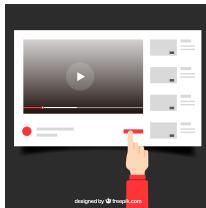
- Always think about what your population is
- Collect data from a sample that is representative of your population
- If you have no choice but to work with a dataset that is not collected randomly and is biased, be careful not to generalize your results to the entire population



You'd want to be sure you sample randomly across *all* YouTube comments, making sure not to get more comments from one genre over another, or one location over another, etc.

Examples of bad sampling:

- Surveying subscribers of a gun-related magazine for research on Americans' attitudes toward owning guns
- Randomly sampling Facebook users for what TV shows people like

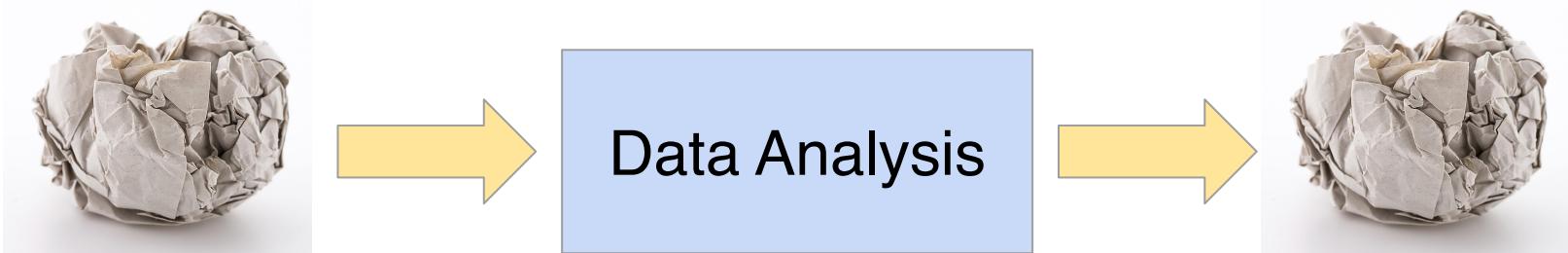


To understand *all* YouTube comments, you wouldn't just want to sample from one YouTube channel, or videos in a single language.

It's *always* worth spending time at the beginning of a project to determine whether or not the data you have are garbage.

Be certain they are actually able to help you answer the question you're interested in.

GIGO : Garbage In. Garbage Out.



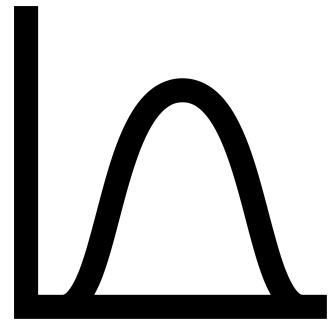
<https://forms.gle/v9bn5kCUVLW5Robc6>

For the survey data I collected from you all, which of the following best describes the population I could generalize findings back to.

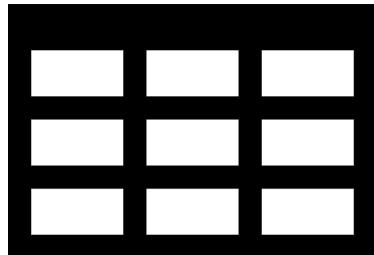
A B C D E



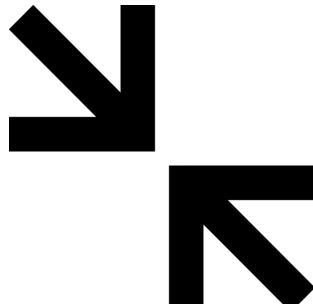
Descriptive Analysis



Shape



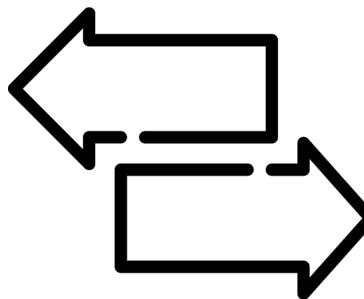
Size



Central
Tendency

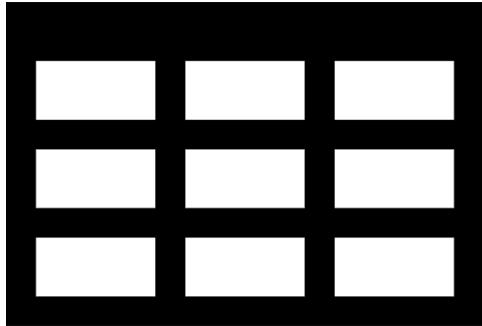


Missingness



Variability

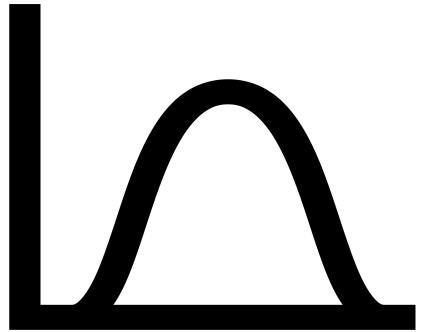
Descriptive



Size

How many observations (rows) and variables (columns) you have is an important first step. You should always be aware of the **size** of your dataset

Missingness It's critical to know how many observations have missing data for variables of interest in your data. Knowing *why* their missing is also important.



Shape

It's critical to know the distribution of the variables in your dataset. Certain statistical approaches can only be used with certain distributions.

Central Tendency

Knowing the mean, median, and/or mode can help you get an idea of what a typical value is for your variable(s) of interest

Variability

The central tendency tells you part of the story. The variability in the values in your observation helps fill in the rest.



Which of the following is NOT something accomplished by a descriptive analysis?

- A** Describes typical values in your dataset
- B** Determines the size of your dataset
- C** Establishes causal relationships between variables
- D** Identifies missing data
- E** Determines how variable values in your dataset are

Descriptive Analyses are often included as “Table 1” in academic publications

Descriptive

Table 1. Baseline Characteristics of the Patients.^a

Characteristic	Ranibizumab Monthly (N=301)	Bevacizumab Monthly (N=286)	Ranibizumab as Needed (N=298)	Bevacizumab as Needed (N=300)
Age — no. (%)				
50–59 yr	2 (0.7)	1 (0.3)	6 (2.0)	2 (0.7)
60–69 yr	33 (11.0)	28 (9.8)	31 (10.4)	34 (11.3)
70–79 yr	102 (33.9)	84 (29.4)	115 (38.6)	103 (34.3)
80–89 yr	142 (47.2)	150 (52.4)	126 (42.3)	142 (47.3)
≥90 yr	22 (7.3)	23 (8.0)	20 (6.7)	19 (6.3)
Mean — yr	79.2±7.4	80.1±7.3	78.4±7.8	79.3±7.6
Sex — no. (%)				
Female	183 (60.8)	180 (62.9)	185 (62.1)	184 (61.3)
Male	118 (39.2)	106 (37.1)	113 (37.9)	116 (38.7)
Race — no. (%) [†]				
White	297 (98.7)	281 (98.3)	296 (99.3)	294 (98.0)
Other	4 (1.3)	5 (1.7)	2 (0.7)	6 (2.0)
History of myocardial infarction — no. (%)	34 (11.3)	40 (14.0)	30 (10.1)	36 (12.0)
History of stroke — no. (%)	14 (4.7)	18 (6.3)	22 (7.4)	16 (5.3)
History of transient ischemic attack — no. (%)	12 (4.0)	25 (8.7)	12 (4.0)	19 (6.3)
Blood pressure — mm Hg				
Systolic	134±18	135±19	136±17	135±17
Diastolic	75±10	75±10	76±9	75±10
Visual-acuity score and Snellen equivalent				
68–82 letters, 20/25–40 — no. (%)	111 (36.9)	94 (32.9)	116 (38.9)	103 (34.3)
53–67 letters, 20/50–80 — no. (%)	98 (32.6)	118 (41.3)	108 (36.2)	119 (39.7)
38–52 letters, 20/100–160 — no. (%)	67 (22.3)	53 (18.5)	58 (19.5)	58 (19.3)
23–37 letters, 20/200–320 — no. (%)	25 (8.3)	21 (7.3)	16 (5.4)	20 (6.7)
Mean score	60.1±14.3	60.2±13.1	61.5±13.2	60.4±13.4
Total thickness at fovea — μm [‡]	458±184	463±196	458±193	461±175
Retinal thickness plus subfoveal-fluid thickness at fovea — μm	251±122	254±121	247±122	252±115
Foveal center involvement — no. (%)				
Choroidal neovascularization	176 (58.5)	153 (53.5)	176 (59.1)	183 (61.0)
Fluid	85 (28.2)	81 (28.3)	77 (25.8)	72 (24.0)
Hemorrhage	20 (6.6)	24 (8.4)	24 (8.1)	25 (8.3)
Other	18 (6.0)	20 (7.0)	15 (5.0)	18 (6.0)
No choroidal neovascularization or not possible to grade	2 (0.7)	8 (2.8)	6 (2.0)	2 (0.7)

* Plus-minus values are means ±SD.

[†] Race was self-reported.

[‡] Total thickness at the fovea includes the retina, subretinal fluid, choroidal neovascularization, and retinal pigment epithelial elevation.

Descriptive

Size

Zooming in on
this we see
variables
stratified by
Age, Sex, and
Race

Table 1. Baseline Characteristics of the Patients.*

Characteristic	Ranibizumab Monthly (N=301)	Bevacizumab Monthly (N=286)	Ranibizumab as Needed (N=298)	Bevacizumab as Needed (N=300)
Age — no. (%)				
50–59 yr	2 (0.7)	1 (0.3)	6 (2.0)	2 (0.7)
60–69 yr	33 (11.0)	28 (9.8)	31 (10.4)	34 (11.3)
70–79 yr	102 (33.9)	84 (29.4)	115 (38.6)	103 (34.3)
80–89 yr	142 (47.2)	150 (52.4)	126 (42.3)	142 (47.3)
≥90 yr	22 (7.3)	23 (8.0)	20 (6.7)	19 (6.3)
Mean — yr	79.2±7.4	80.1±7.3	78.4±7.8	79.3±7.6
Sex — no. (%)				
Female	183 (60.8)	180 (62.9)	185 (62.1)	184 (61.3)
Male	118 (39.2)	106 (37.1)	113 (37.9)	116 (38.7)
Race — no. (%)†				
White	297 (98.7)	281 (98.3)	296 (99.3)	294 (98.0)
Other	4 (1.3)	5 (1.7)	2 (0.7)	6 (2.0)

* Plus-minus values are means ±SD.

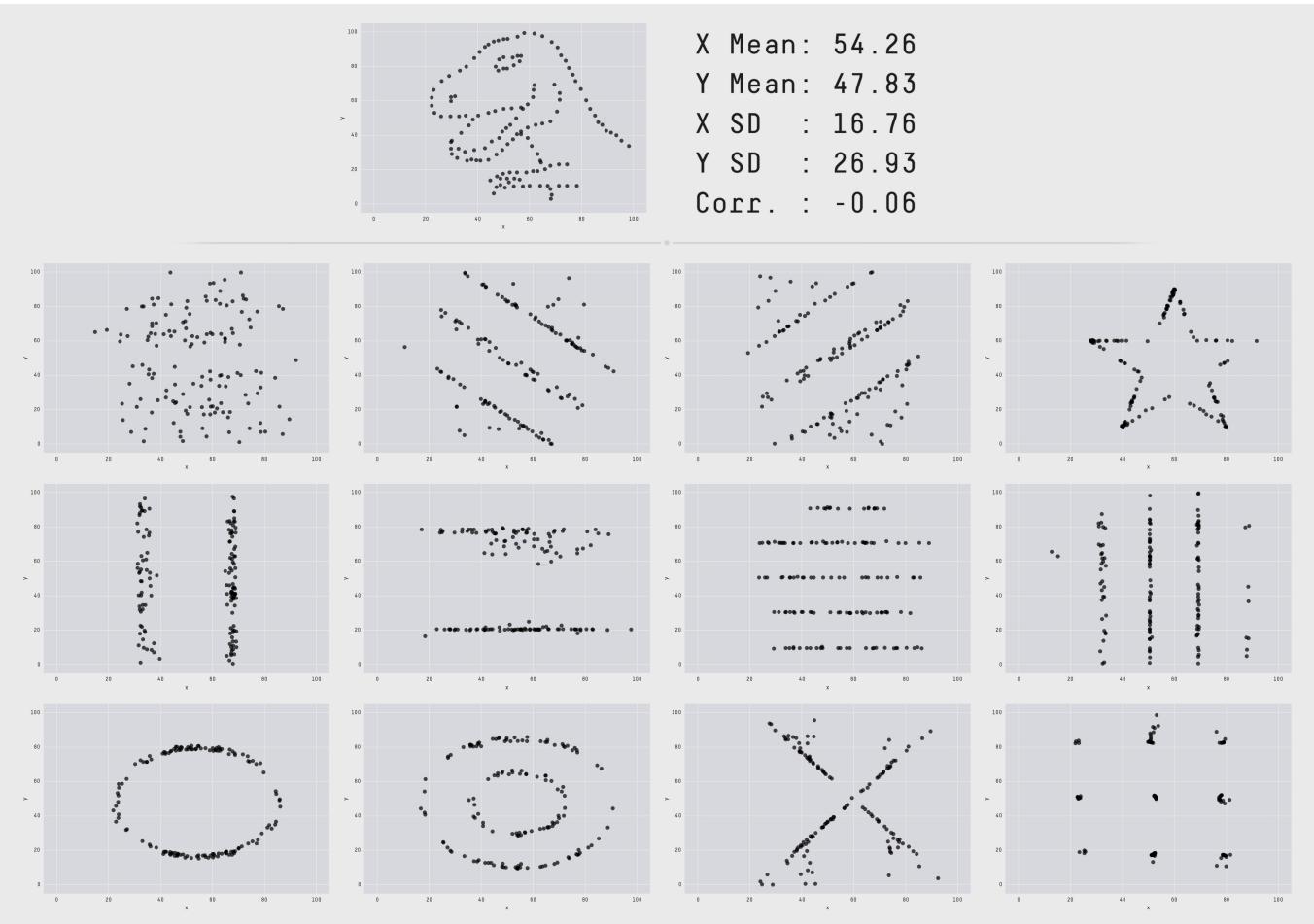
† Race was self-reported.

‡ Total thickness at the fovea includes the retina, subretinal fluid, choroidal neovascularization, and retinal pigment epithelial elevation.

Shape
Central
tendency
variability

Course Reminders

- Due Wed (11:59PM)
 - Project Review
- Due Friday (11:59 PM)
 - D3
 - Weekly Project Survey (*optional EC; released on Wed*)
 - Github invite expired? Ask on ED... there's a thread!



Descriptive Statistics & Summary

“We must suppress some of the truth to communicate the truth... In short, the techniques of descriptive statistics are designed to match the salient features of the data set to human cognitive abilities.”

-I.J. Good (1983)

Descriptive

Descriptive Statistics & Summary

Calculating descriptive statistics, understanding what they tell you about your data, and reporting them are critical steps in every analysis.

Yes statistics are summaries that throw away important detail, but human minds need the high level overview since we often struggle with the details.

Exploratory

Exploratory: The goal is to find unknown relationships between the variables you have measured in your data set. Exploratory analysis is open ended and designed to verify expected or find unexpected relationships between measurements.

Exploratory



Exploratory Data Analysis (EDA)
detective work answering the question:
“What can the data tell us?”

Exploratory

What do we get from EDA?

- Understand data properties
- Discover Patterns
- Generate & Frame Hypothesis
- Suggest modeling strategies
- Check assumptions (sanity checks)
- Communicate results (present the data)

Exploratory

The
dataset

You



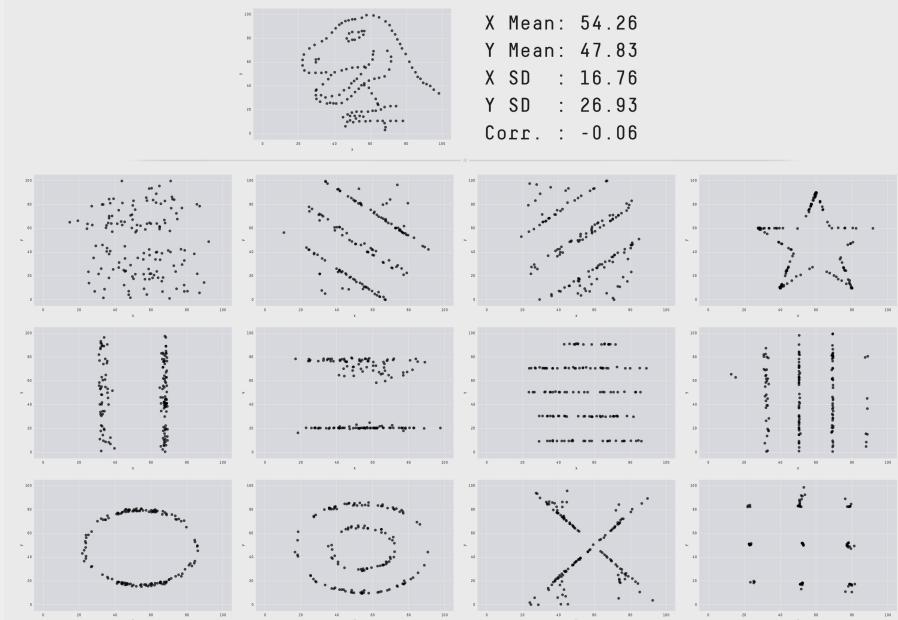
Exploratory

The general principles of exploratory analysis:

- Look for missing values
- Look for outlier values
- Calculate numerical summaries
- Generate plots to explore relationships
- Use tables to explore relationships
- If necessary, transform variables

Start raw

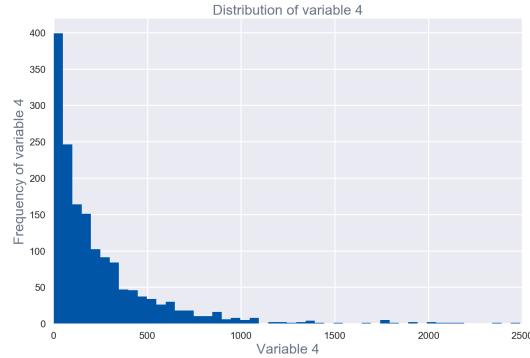
- Examine raw data in the most direct way you can reasonably do so
- View a random sample of the data
- Plots, especially subsets of variables and dimensionality reduction
- Helpful for seeing weirdness, missingness, outliers, min/max/typical values



Exploratory

EDA Approaches to “Get a Feel for the Data”

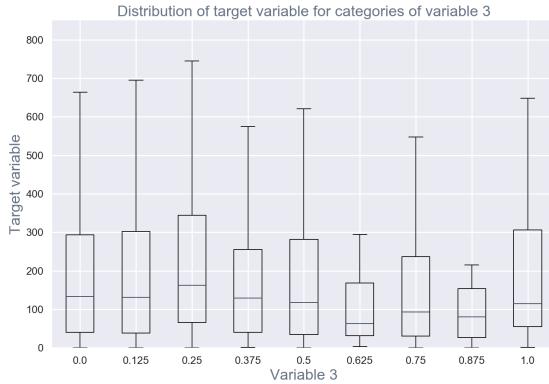
Understanding the relationship between variables in your dataset



Univariate

understanding a single variable

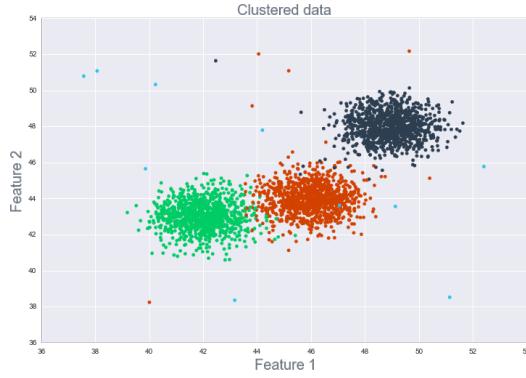
i.e.: histogram, densityplot, barplot



Bivariate

understanding relationship between 2 variables

i.e.: boxplot, scatterplot, grouped barplot, boxplot

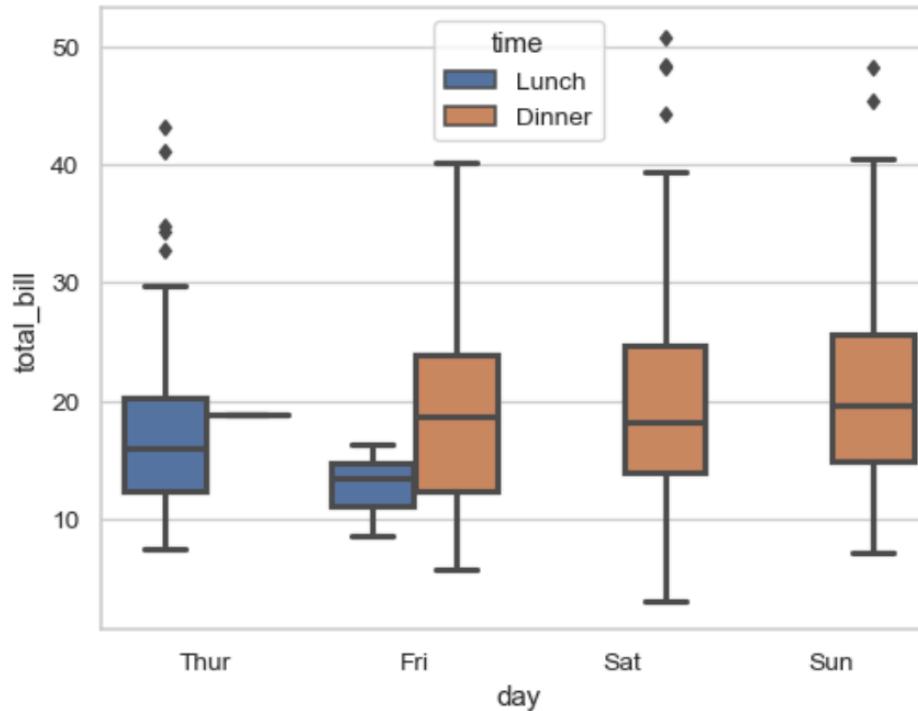


Dimensionality Reduction

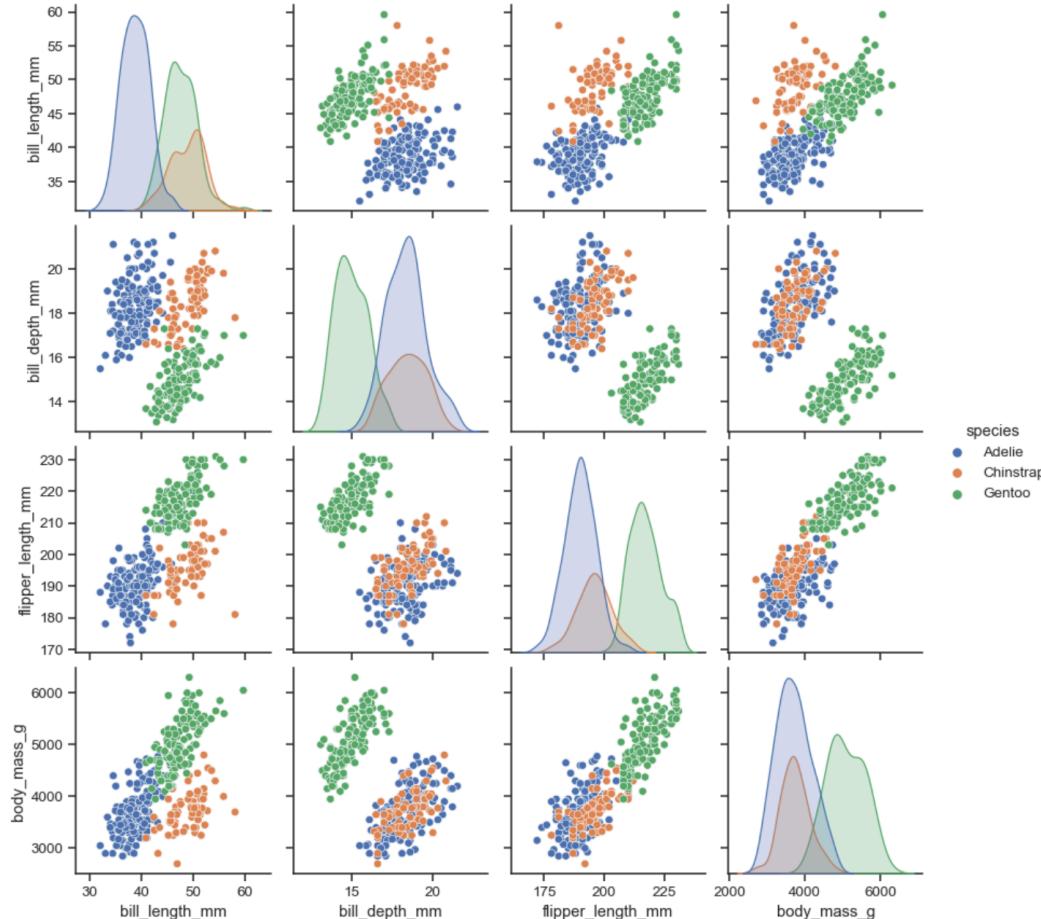
projecting high-D data into a lower-D space

i.e.: PCA, ICA, Clustering

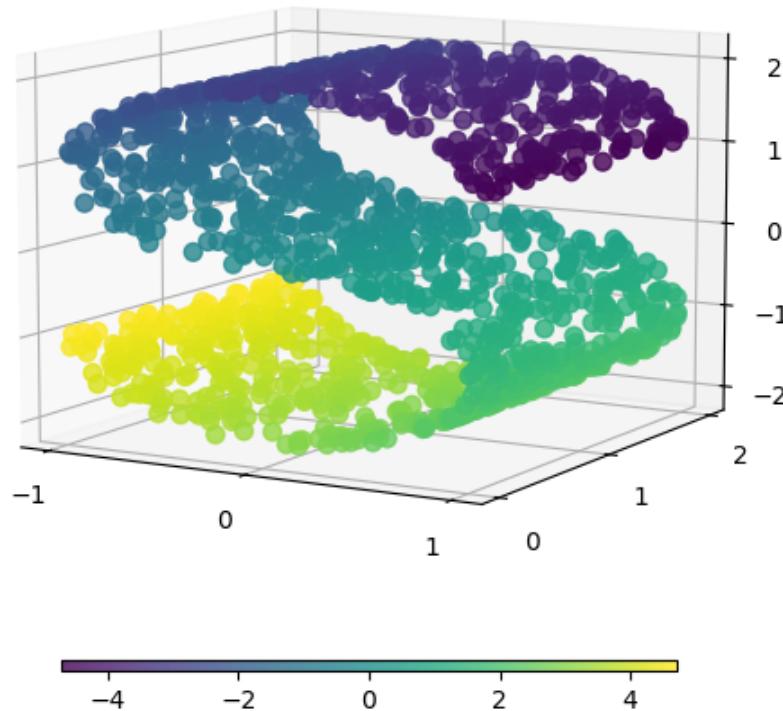
```
>>> ax = sns.boxplot(x="day", y="total_bill", hue="time",
...                     data=tips, linewidth=2.5)
```

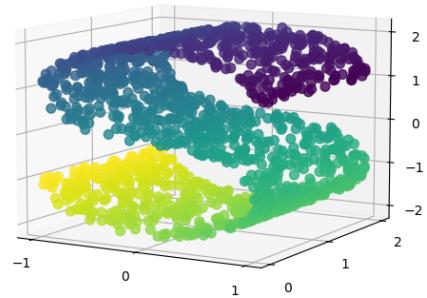


```
sns.pairplot(penguins, hue="species")
```

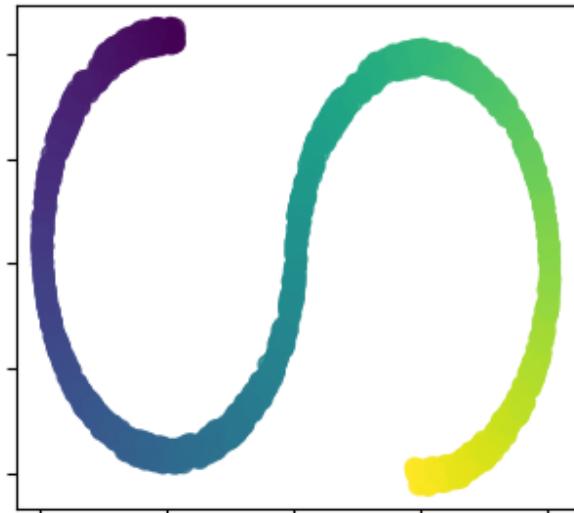


Original S-curve samples

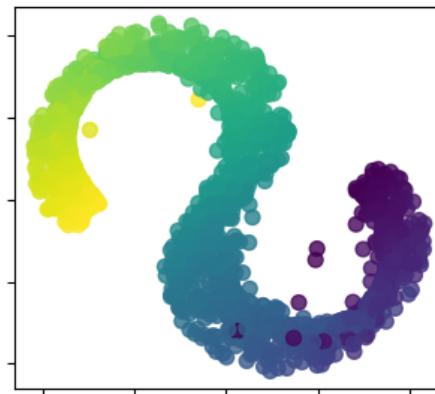




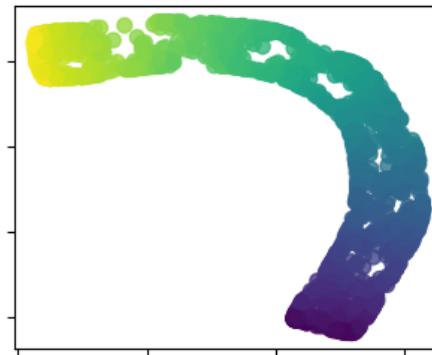
PCA



Multidimensional scaling



T-distributed Stochastic
Neighbor Embedding



MNIST data T-SNE projection

