

Video Project #2 Topics

The following are the topics for the Video Project. The number you received in the PDF posted for your Discussion section corresponds to the Project you were randomly selected to do.

In addition to the topic itself, we list a few ideas for getting started. Each of these topics relates quite closely to concepts from the course lectures and in most cases your textbook and supplemental videos. You are encouraged to review all course material related to these topics.

Of course, each of these topics are important beyond our specific course! Indeed, most of these topics have excellent Wikipedia articles. You are heavily encouraged to search for additional resources on-line.

Remember, there are no “right answers” we have in mind for your video projects. We encourage you to be creative and have fun!

You are encouraged to consult with the course Instructors/TAs and your classmates on this video topic. However, each student is responsible for submitting their own video project. **Finally**, remember your video project is due at the same time as the scheduled final exam ends (**December 17th by 9:30pm**).

Category 1: Data Visualization

Topic #	Topic Name	Ideas for Getting Started
1	Kernel Density Plots (Triangular Kernel)	What are kernel density plots? What is a the triangular kernel? Why is this useful? Show some example data.
2	Kernel Density Plots (Cosine Kernel)	What are kernel density plots? What is the cosine kernel? Why is this useful? Show some example data.
3	Histograms	What does this mean? What kinds of options do you have with this type of visualization? Why is this useful? Show some example data.
4	Box & Whisker Plots	What does this mean? What kinds of options do you have with this type of visualization? What about outliers? Why is this useful? Show some example data.
5	Stem & Leaf Plots	What does this mean? What kinds of options do you have with this type of visualization? When is this a useful way to present data? Show some example data.
6	Violin Plot	What is this type of visualization? What kinds of options do you have with this type of visualization? When is it useful?
7	Jitter Plots	What is this type of visualization? What kinds of options do you have with this type of visualization? When is it useful?

Category 2: Important Statistical Concepts

Topic #	Topic Name	Ideas for Getting Started
8	Mean vs Median	What do each of these mean? When are they the same? When are they different?
9	Mean vs Sample Mean	What do each of these mean? When are they the same? When are they different?
10	Variance vs Sample Variance (Unbiased Estimator for)	What do each of these mean? Why are they different?
11	Median Absolute Deviation vs Variance	How are these related? When are they similar? How are each impacted by outliers? Why are they each useful? Give examples.
12	Correlation	What does this mean? What does it tell us? Can it mis-lead us?
13	Linear Regression	When do you use this? How does it work? What does it tell us?
14	Confidence Intervals (Mean of Normal Distribution)	What is a confidence interval? How do you calculate a confidence interval in the mean of a normal distribution? Give an example.
15	Confidence Intervals (Probability of Success in Binomial Distribution)	What is a confidence interval? How do you calculate a confidence interval in the success probability of a binomial distribution? Give an example.
16	Law of Large Numbers	What does it mean mathematically? What does it mean practically? Give examples.
17	Chebyshev's Inequality	What does it mean mathematically? What does it mean conceptually? Give examples.
18	Central Limit Theorem	What does it mean mathematically? What does it mean practically/conceptually? Give examples of where it's used.
19	Estimators (Biased vs Unbiased)	What is an estimator? What properties should it have? What does it mean for an estimator to be unbiased? Give examples.
20	Maximum Likelihood Estimation	What does it mean mathematically? What does it mean conceptually? Give examples.
21	Hypothesis Testing	What is it? How do you use it? Why is it useful?
22	p -value	What is it? What does it mean? How does it relate to hypothesis testing?
23	Empirical Bootstrap	What is it? What does it mean? When would you use it? (Check out Chapter 18 in your textbook.)

Category 3: Statistical Tests

Topic #	Topic Name	Ideas for Getting Started
24	One-Sample t -test	What is it? When do you use it? Give an example.
25	Two-Sample t -test (equal variance)	What does this mean? When do you use it? Give an example.
26	Paired samples t -test	What does this mean? When do you use it. Give an example.

Category 3: Classical Statistics Problems/Puzzles

27	Simpson's Paradox	A classical problem where unnoticed differences in subgroups can cause massive differences in data set behavior. https://en.wikipedia.org/wiki/Simpson%27s_paradox
28	Survivorship Bias and Abraham Wald's Analysis of Planes	How do you tell which parts of a plane are the most important to protect? The answer Abraham Wald (https://en.wikipedia.org/wiki/Abraham_Wald) found was counter intuitive because of survivorship bias https://en.wikipedia.org/wiki/Survivorship_bias
29	German Tank Problem	A real world story from WWII where researchers wanted to estimate the maximum of a discrete uniform distribution from sampling without replacement. https://en.wikipedia.org/wiki/German_tank_problem
30	Estimating the Number of Members in a Population (Capture/Recapture)	How do ecologists determine the number of animals in a population? They tag and monitor individuals and use statistics! https://en.wikipedia.org/wiki/Mark_and_recapture
31	Correlation vs Causation	It is commonly said "Correlation does not imply causation." Yet this mistake is made many times. Discuss what this means in the context of this course and give some examples. https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation .