

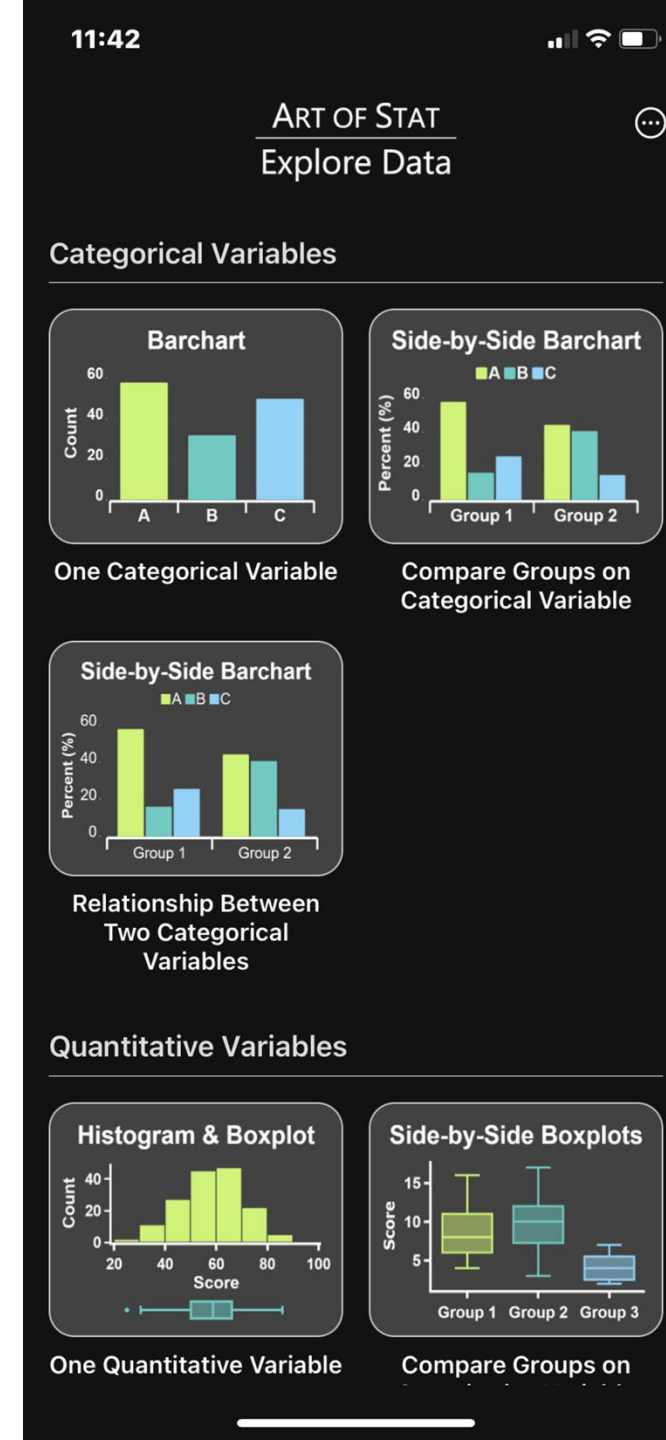
Art of Stat

Mobile Apps for Intro Stats

Bernhard Klingenberg

Prof. of Statistics

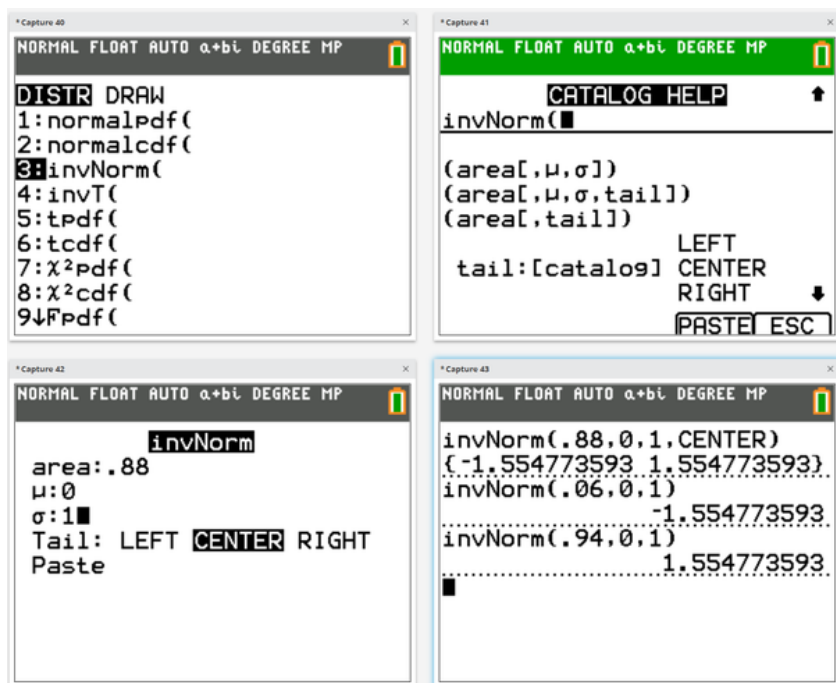
Williams College



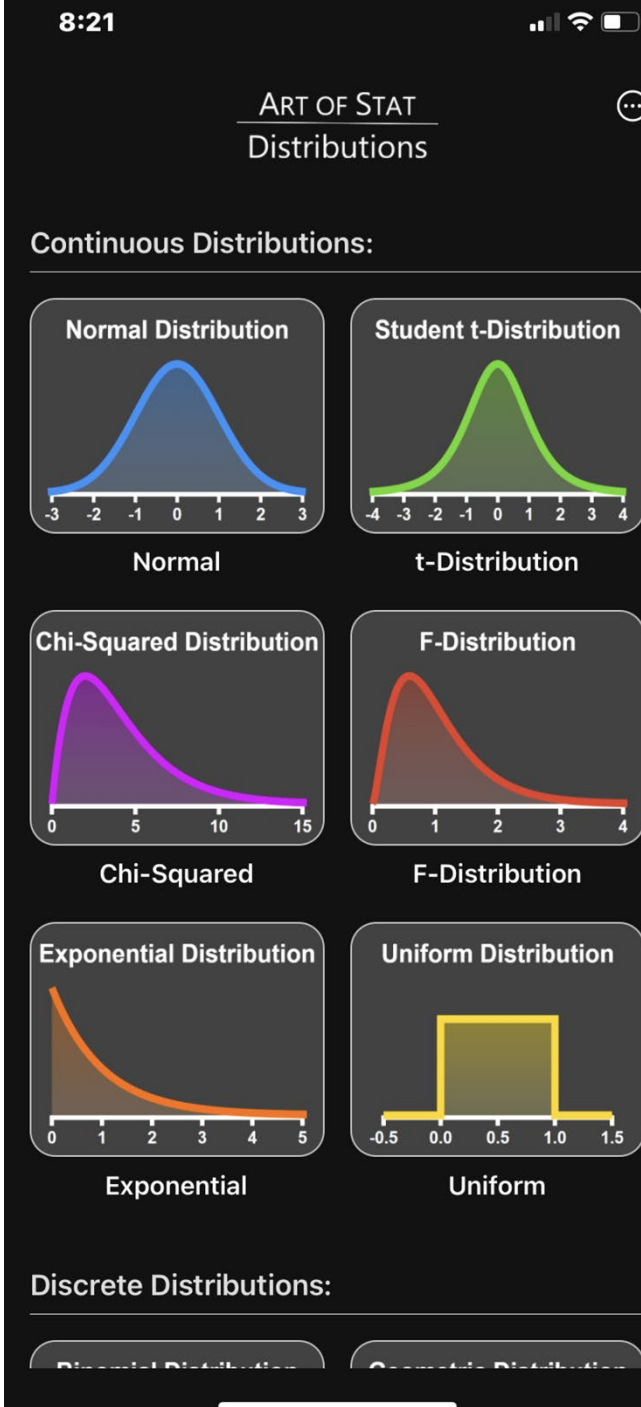
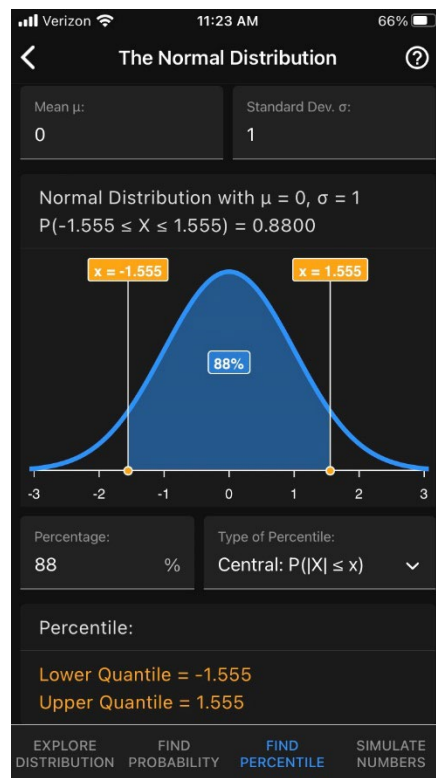
Overview

Conference Theme: “Preparing the Modern Student”

The modern student deserves a modern calculator!

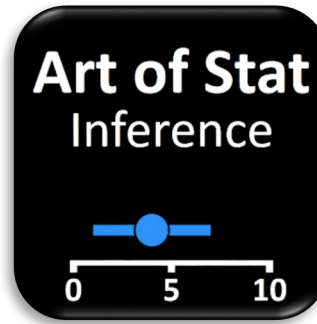
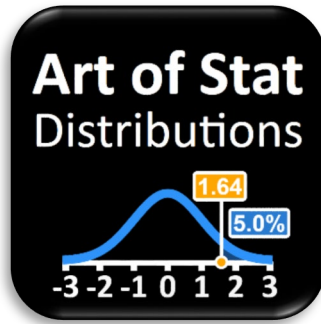
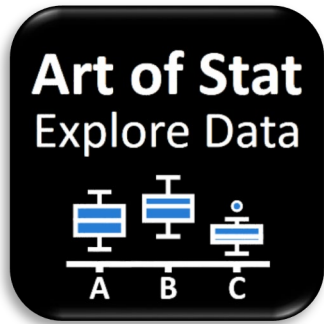


VS.

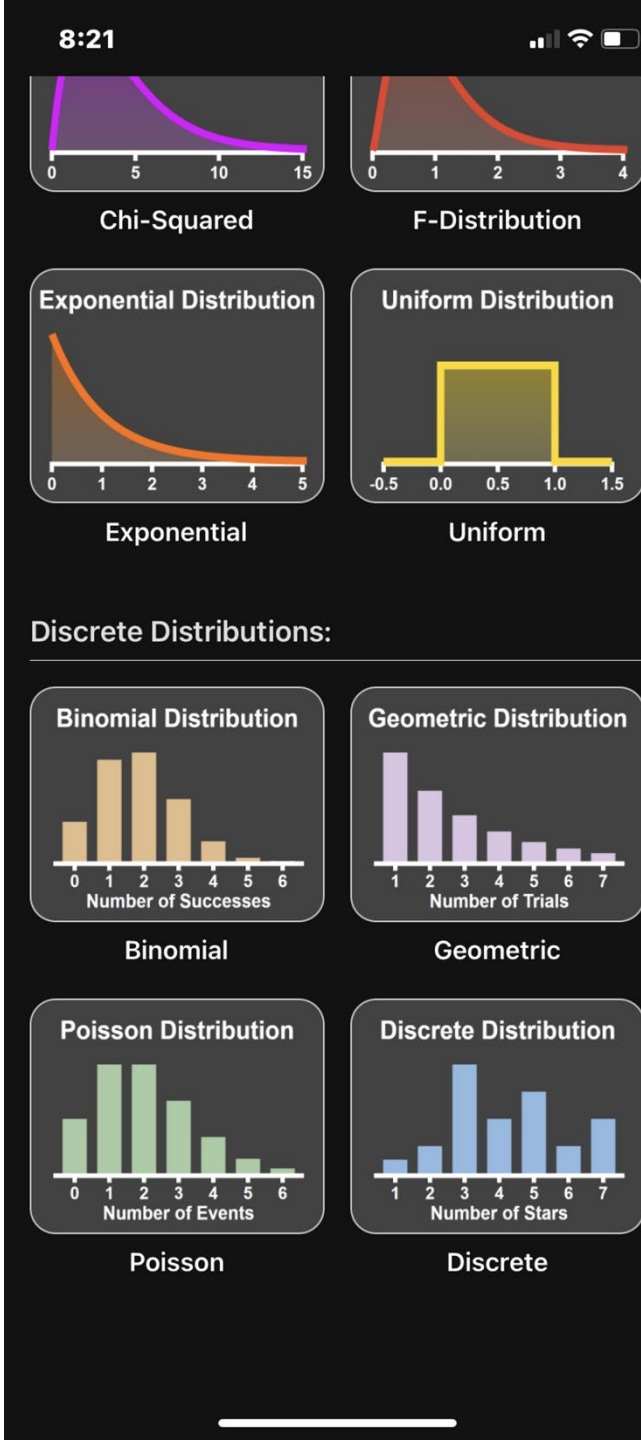


Overview

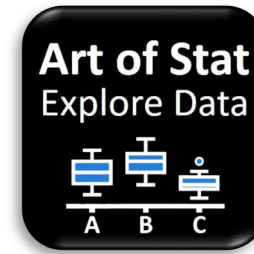
- Apps span four major themes:



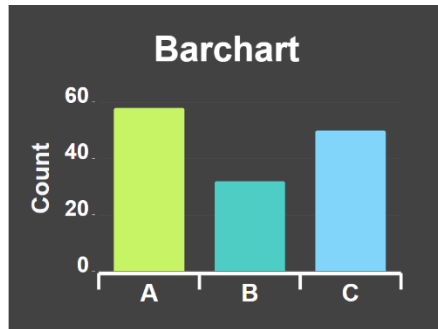
- Available for iOS and Android (search for “Art of Stat”)



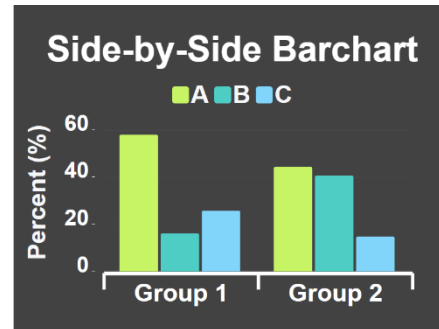
Overview: Explore Data



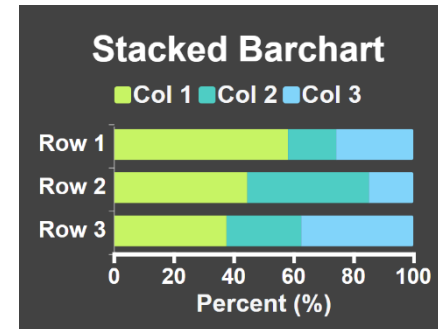
■ Explore Categorical Variables



Distribution of One Variable

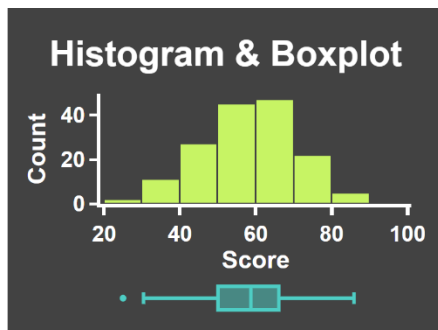


Comparing Groups

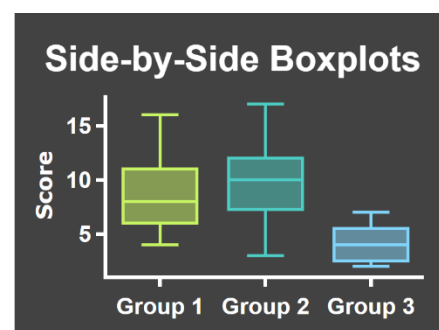


Relationship Btw. Two Variables

■ Explore Quantitative Variables



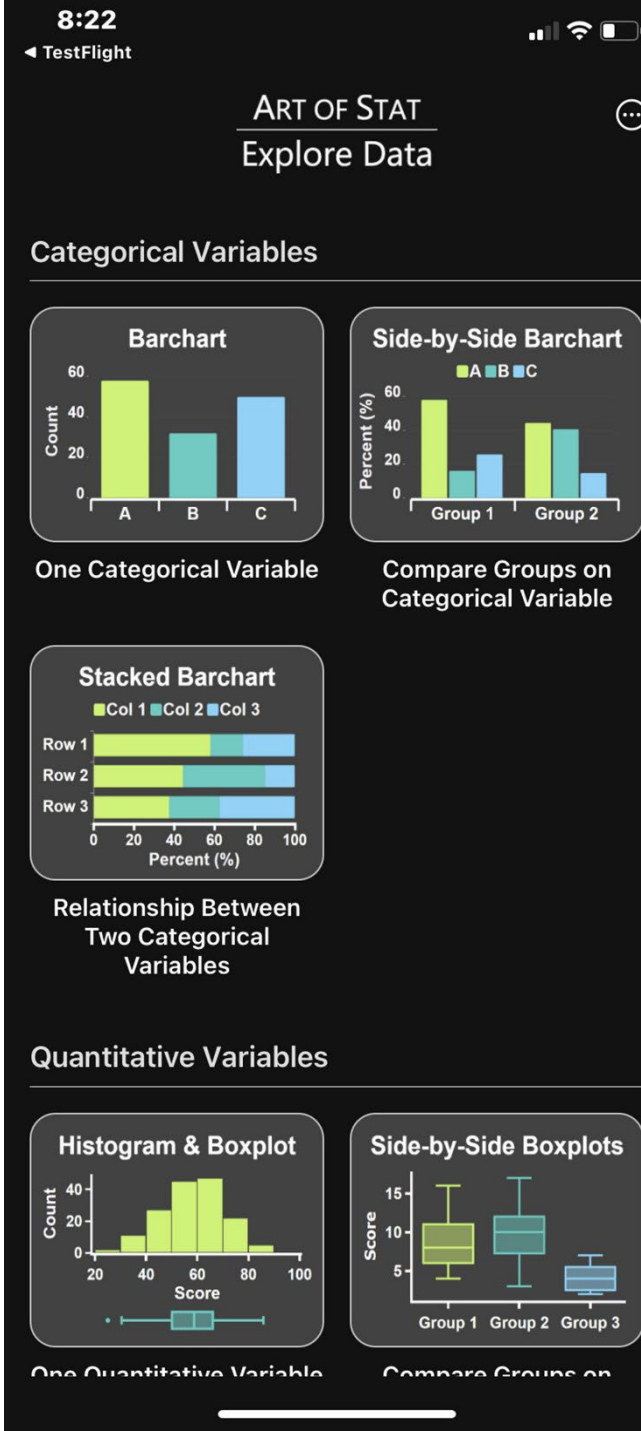
Distribution of One Variable



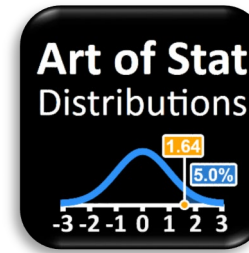
Comparing Groups



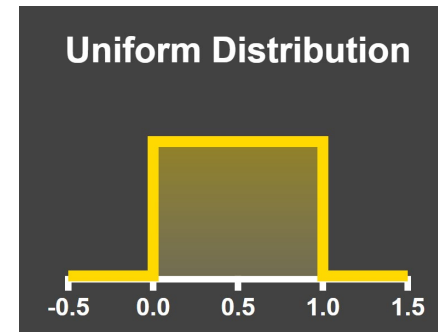
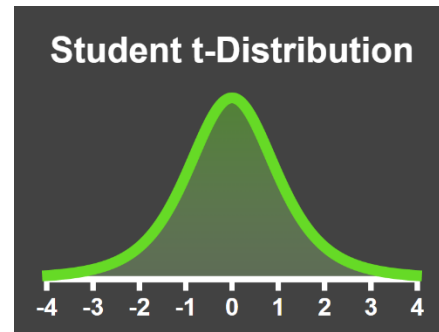
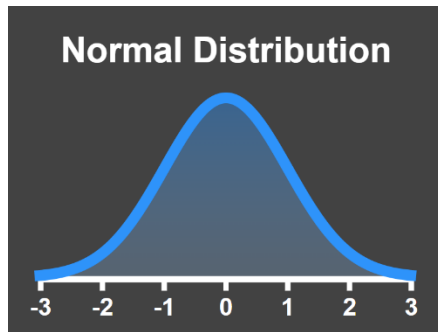
Relationship Btw. Two Variables



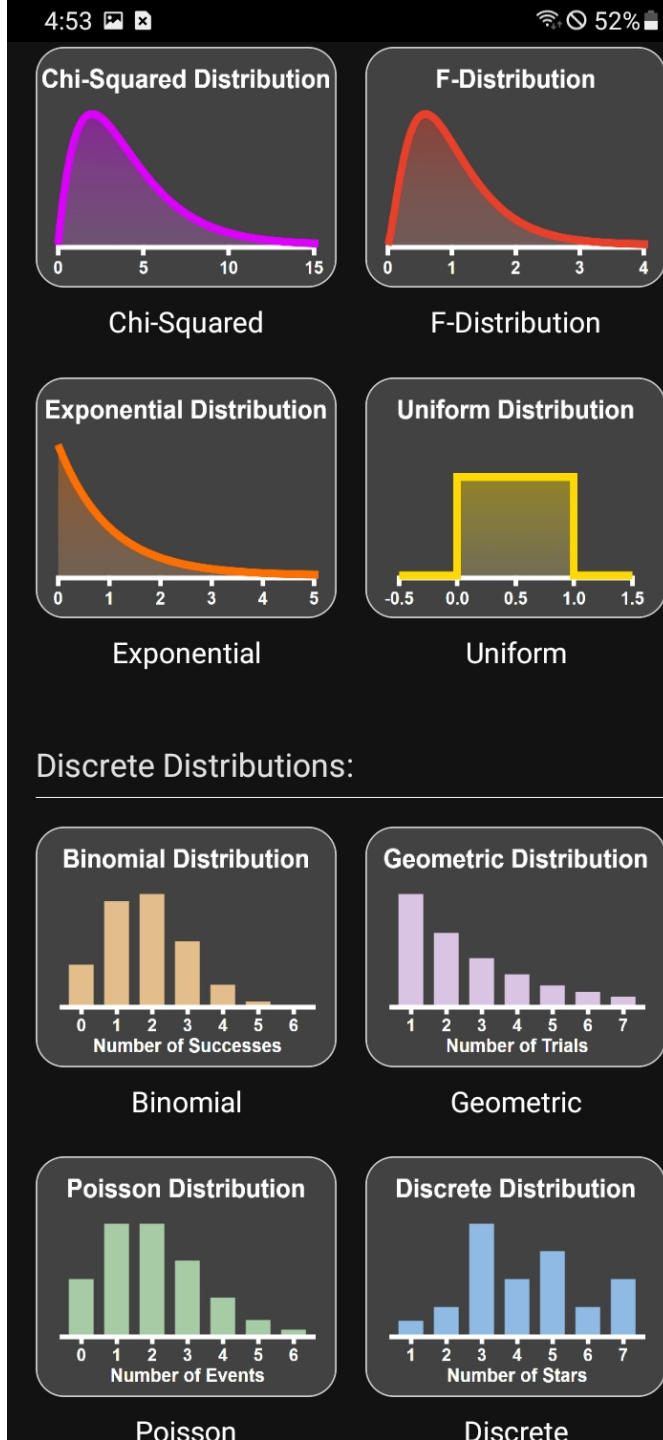
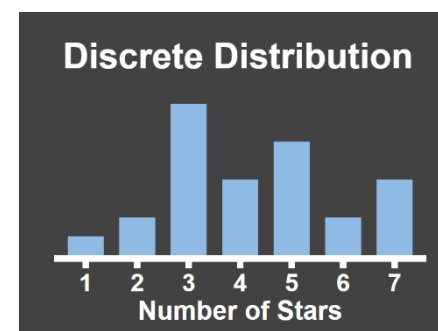
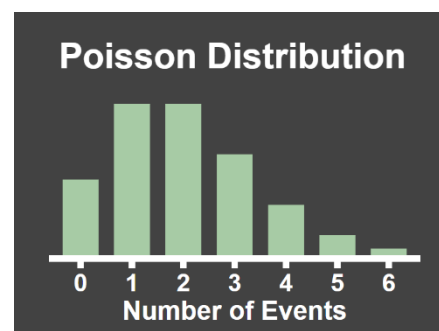
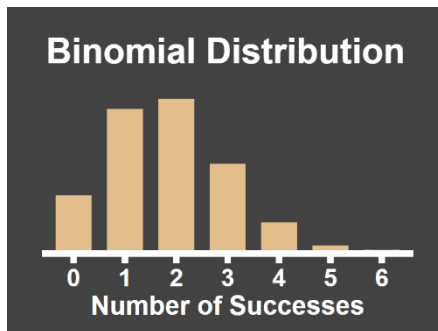
Overview: Distributions



Continuous Distributions

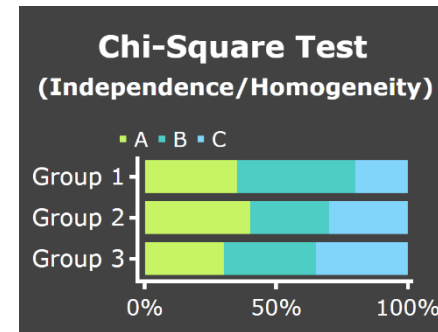
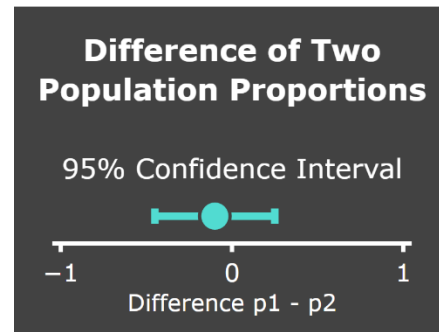
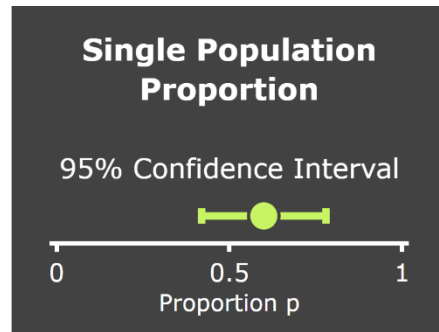


Discrete Distributions

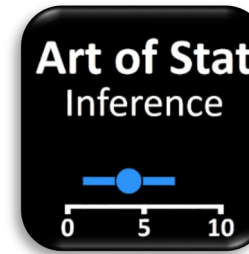
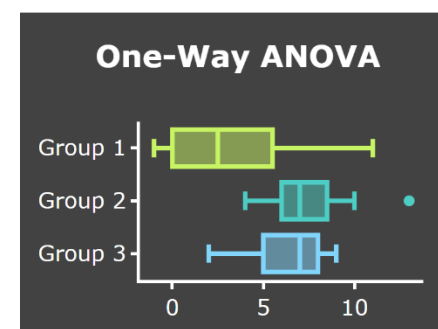
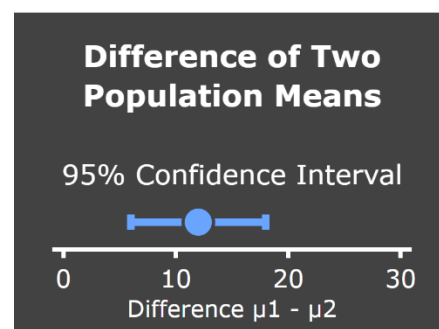
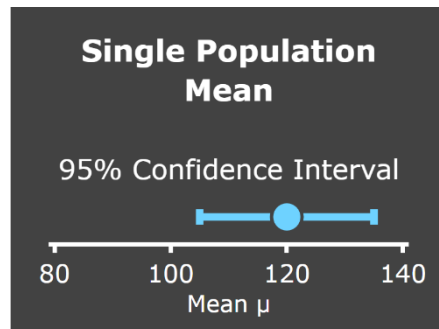


Overview: Inference

■ For Population Proportions



■ For Population Means



8:48

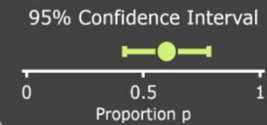


ART OF STAT Inference



Inference About Proportions

Single Population Proportion



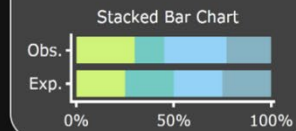
Difference of Two Population Proportions



Inference for a Population Proportion (One Sample)

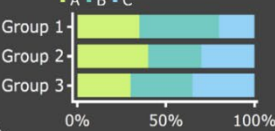
Compare Two Population Proportions (Independent Samples)

Chi-Squared Test (Goodness of Fit)



Chi-Squared Test (Goodness of Fit)

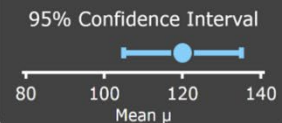
Chi-Squared Test (Independence/Homogeneity)



Chi-Squared Test (Independence or Homogeneity)

Inference About Means

Single Population Mean

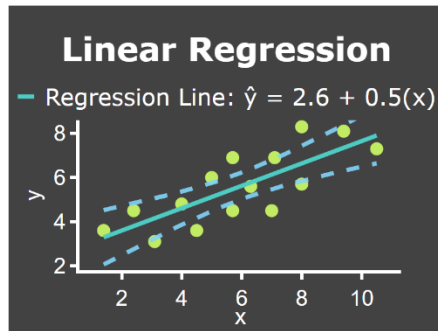


Difference of Two Population Means

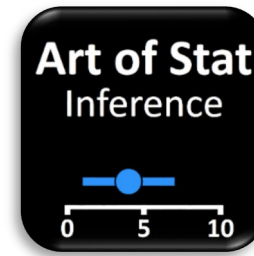
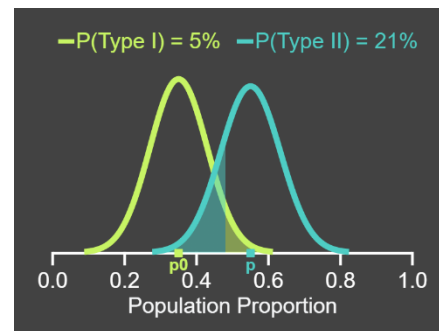
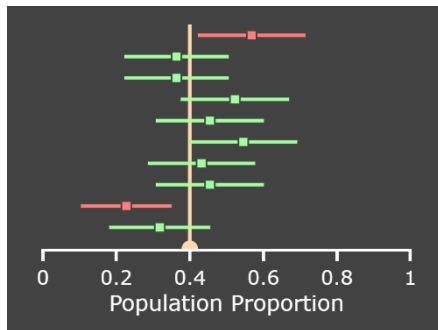


Overview: Inference

■ Inference in Linear Regression



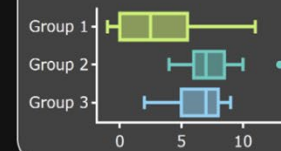
■ Illustrating Concepts: Coverage and Errors & Power



8:49



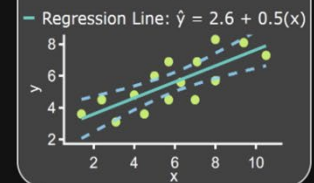
One-Way ANOVA



One-Way ANOVA

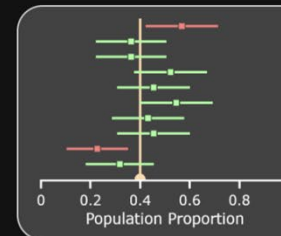
Inference for Linear Regression Model

Linear Regression

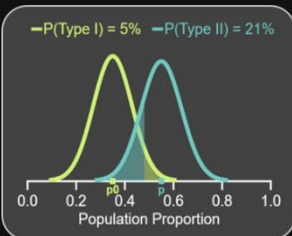


Inference in Linear Regression

Illustrating Concepts



Explore Coverage of Confidence Intervals



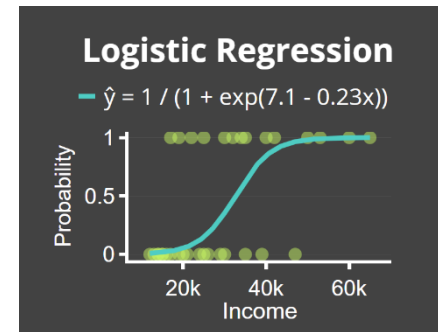
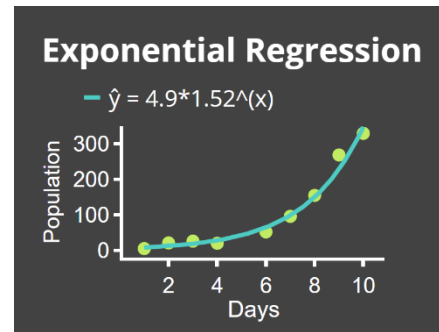
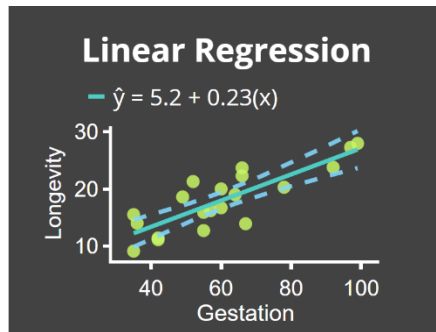
Explore Type I & II Errors and Power

Check Out Other Apps

Overview: Regression

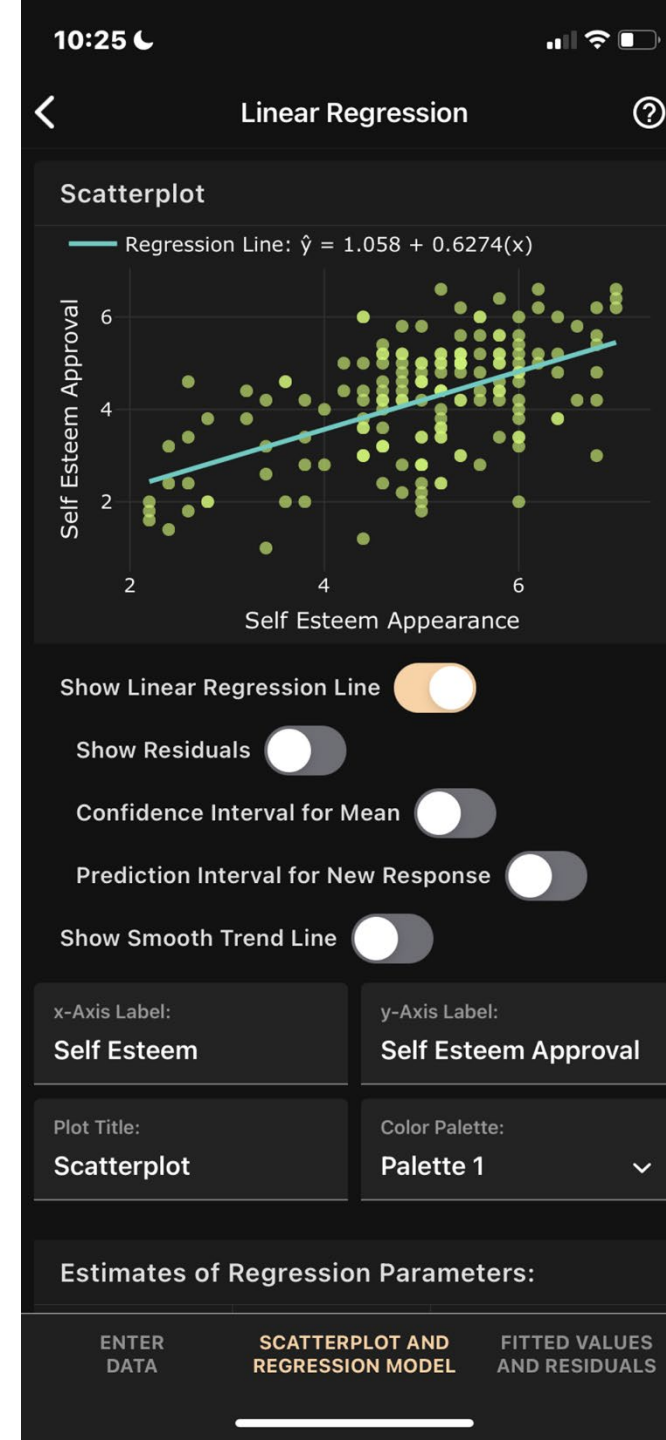


- Simple Regression



- Multiple Regression (coming fall '22)

- Multiple Linear Regression
- Multiple Logistic Regression





Let's try it with proportion inference

Elon Musk ✓ @elonmusk · May 13 ..
Twitter deal temporarily on hold pending details supporting calculation that spam/fake accounts do indeed represent less than 5% of users .
To find out, my team will do a random sample of 100 followers of @twitter.

I invite others to repeat the same process and see what they discover ...

Pranay Pathole @PPathole · May 13 ...
Replying to @elonmusk and @Twitter
Elon, can you elaborate a bit on the "process"? So that we as Twitter can help you in finding out the real percentage of scam/spam/bot accounts

Elon Musk ✓ @elonmusk · May 13 ...
Any sensible random sampling process is fine. If many people independently get similar results for % of fake/spam/duplicate accounts, that will be telling.

I picked 100 as the sample size number, because that is what Twitter uses to calculate <5% fake/spam/duplicate.

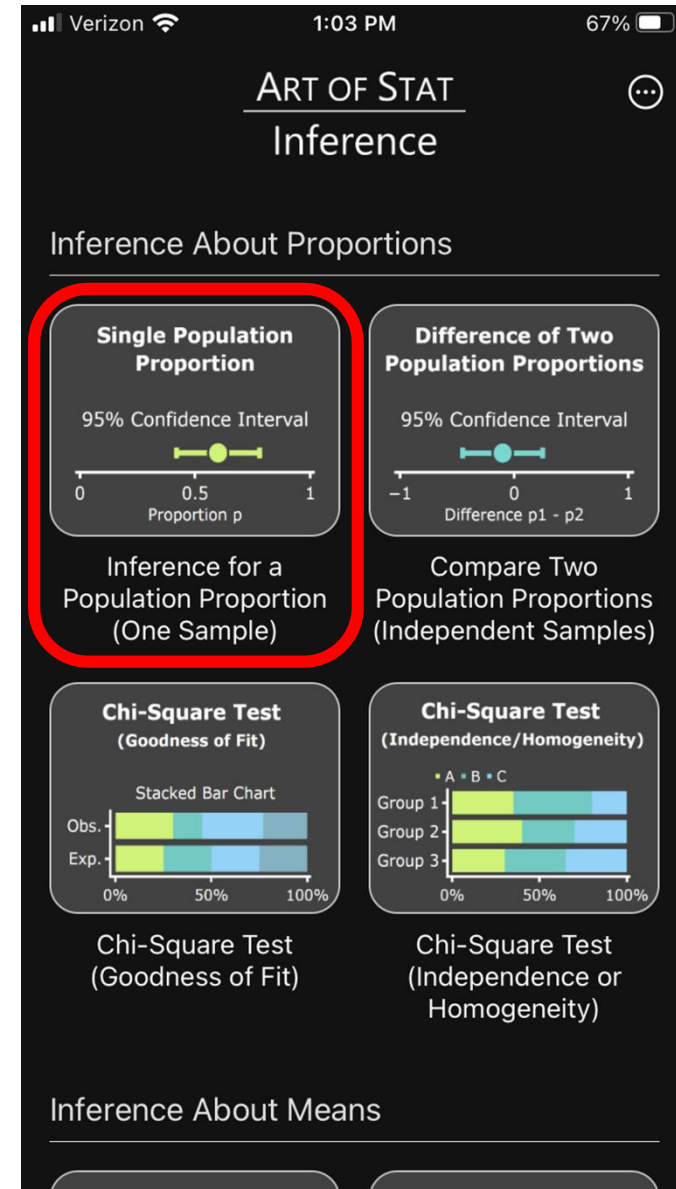
Tesla Owners Silicon... @teslaowne... · May 13 ...
Replying to @elonmusk and @Twitter
How are defining random ? How will you select them

Elon Musk ✓ @elonmusk · May 13 ...
Ignore first 1000 followers, then pick every 10th. I'm open to better ideas.

Coffee Table Tesla (☕) ... @coffeet... · May 13 ...
Replying to @elonmusk and @Twitter
Why followers of that account, specifically? Why not yours?

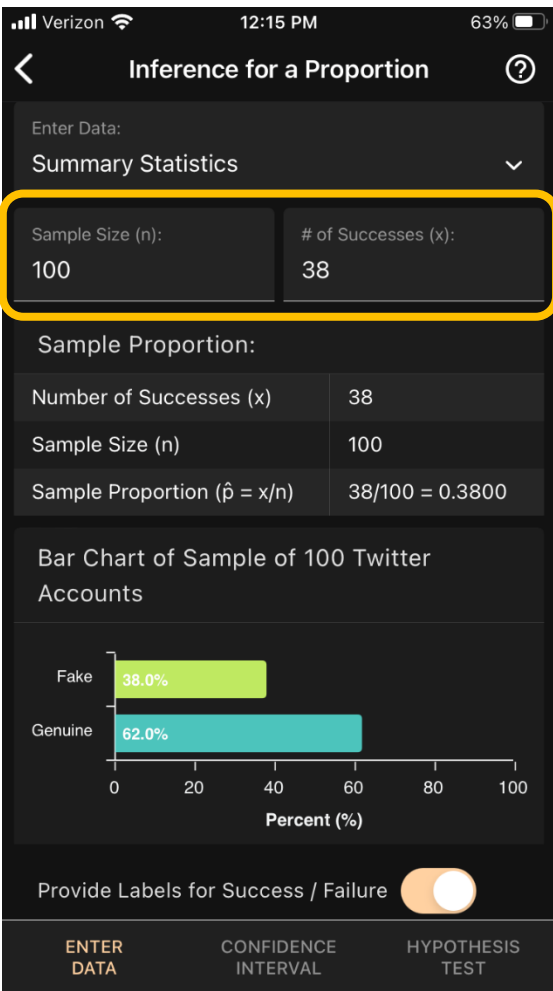
Elon Musk ✓ @elonmusk · May 13
Pick any account with a lot of followers

ໂຮ້ໂຮ້ @realworldNeo · May 14
Replying to @elonmusk and @Twitter
I did @nytimes and it had 38 out of 100

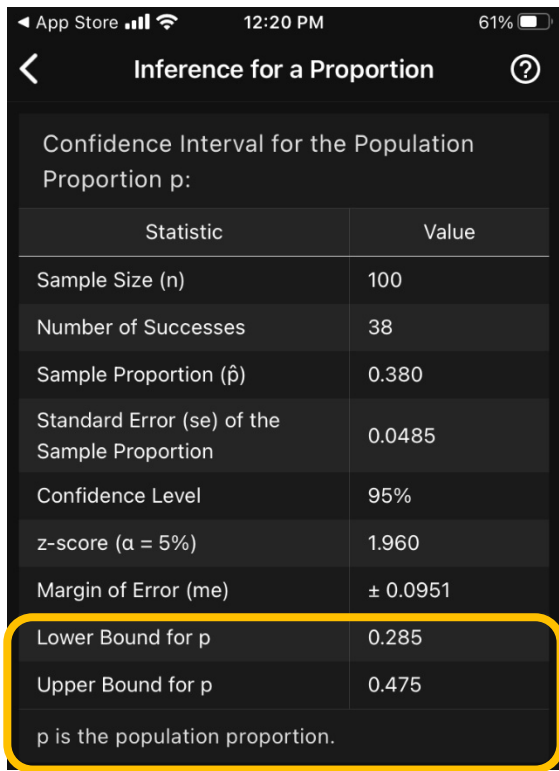




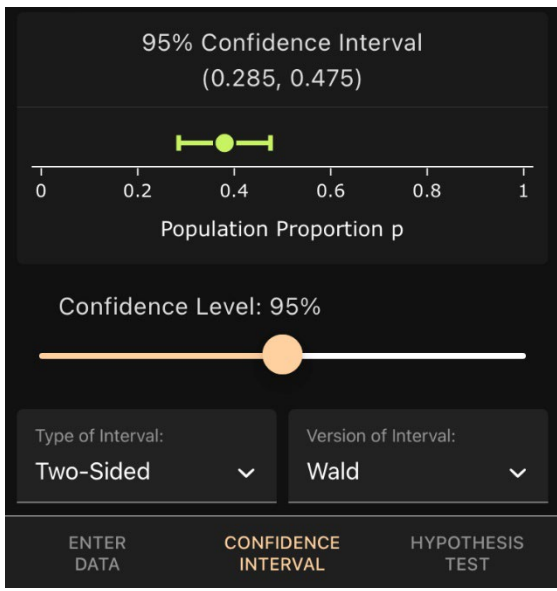
Let's try it with proportion inference



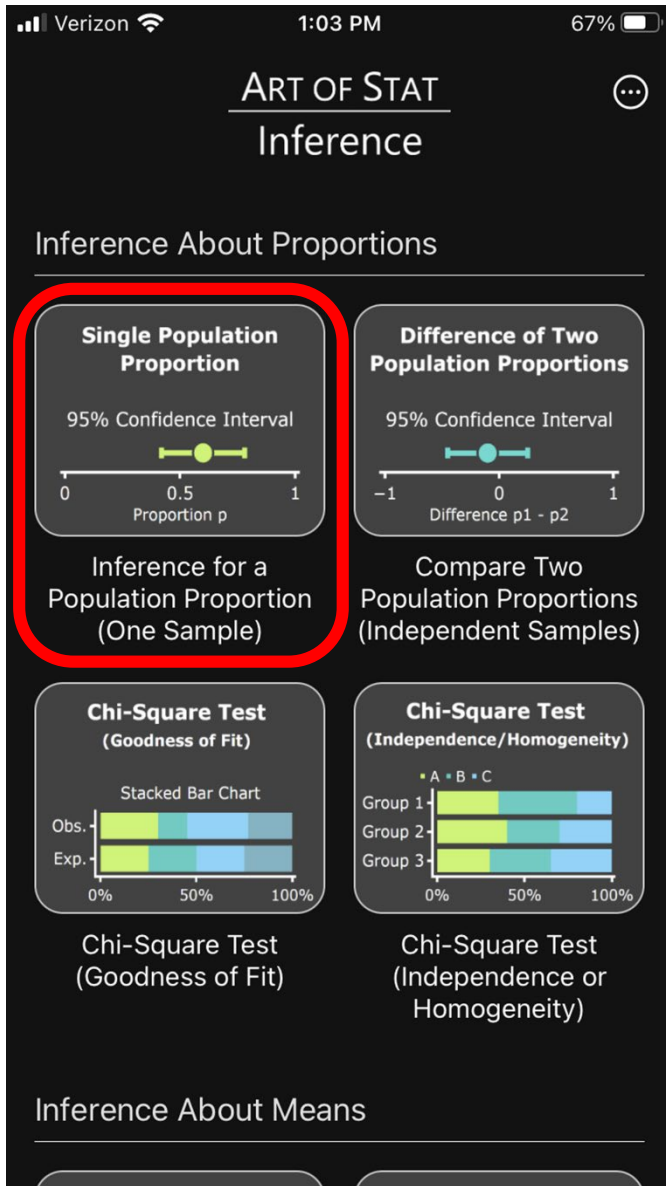
Enter Data on First Tab



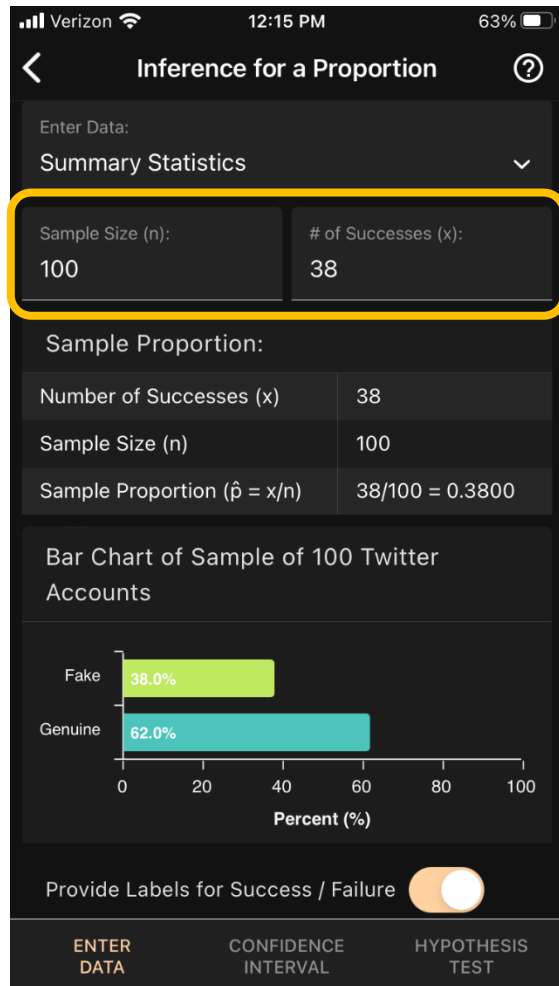
Go to Second Tab
"Confidence Interval" and
obtain lower and upper
bounds (plus intermediate
statistics such as ME)



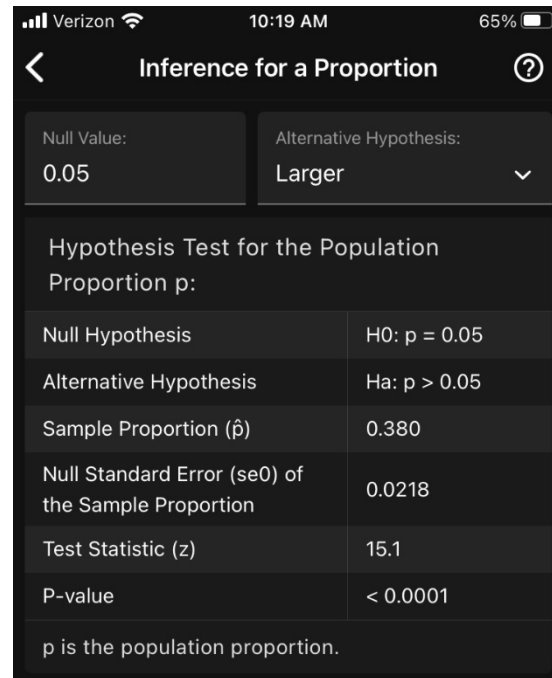
Get a graphical representation of the interval, plus options such as choosing the confidence level



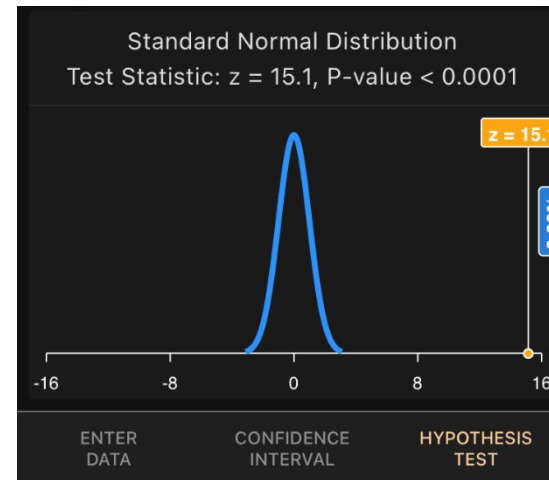
Let's try it with proportion inference



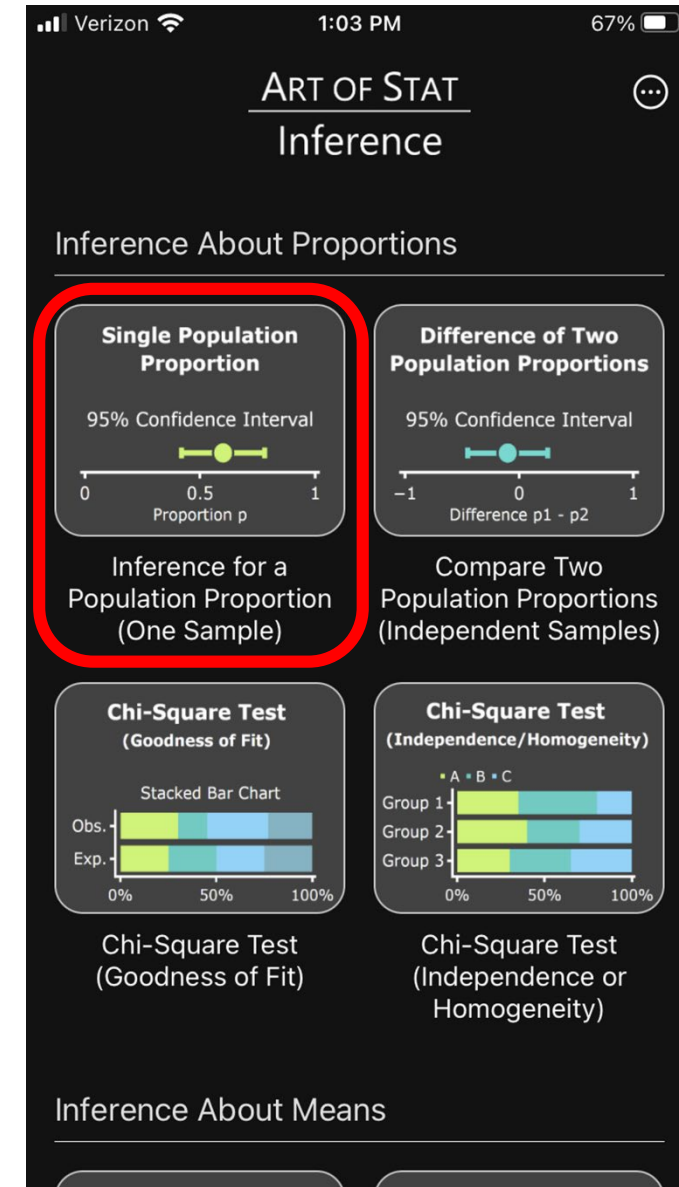
Enter Data on First Tab



Go to Third Tab
“Hypothesis Test”, provide null value and type of hypothesis, and obtain P-value and intermediate steps.

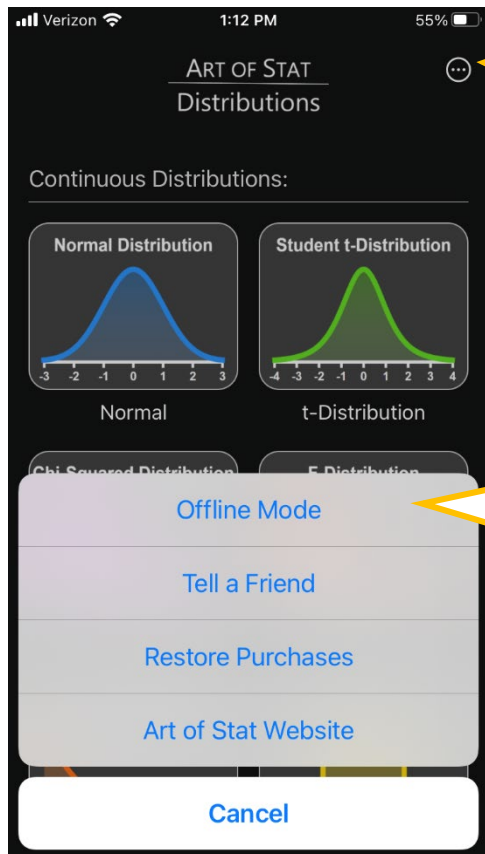


Get a graphical representation of the distribution of the test statistic under the null, and the P-value.



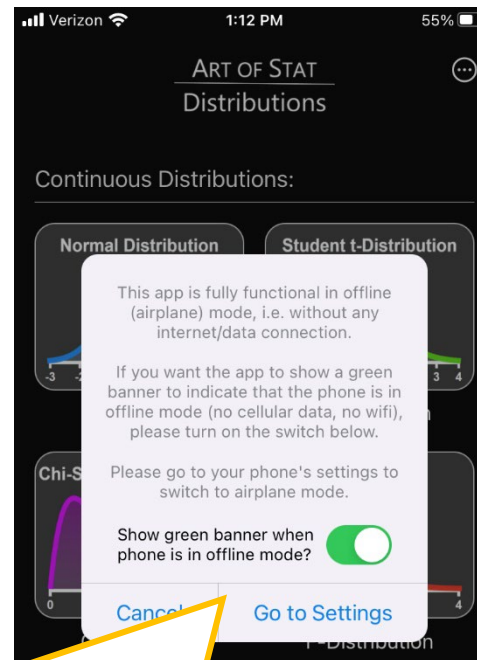
Teaching with the Apps

- The apps work in offline mode:



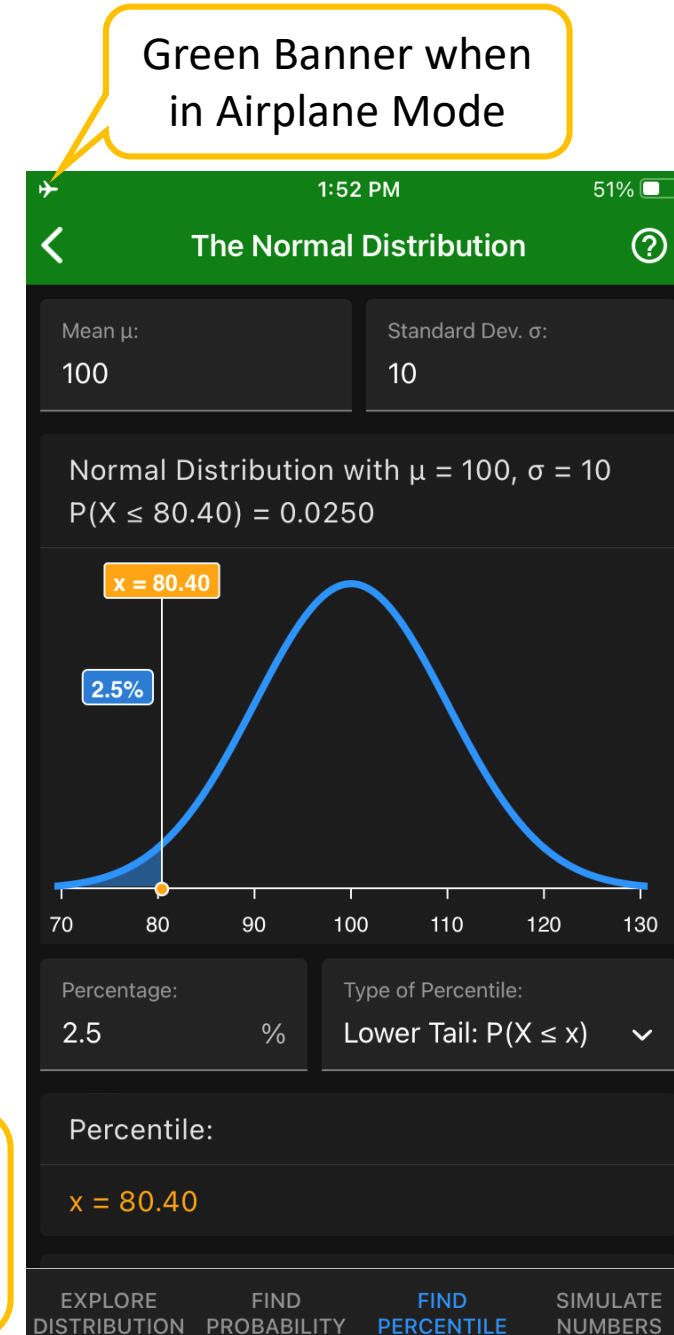
On initial page of an app, tab the three dots in the upper right corner

Select "Offline Mode" from menu



Decide if you want to display green banner if phone is in offline mode (for exams?)

Then, go to settings to turn on Airplane Mode

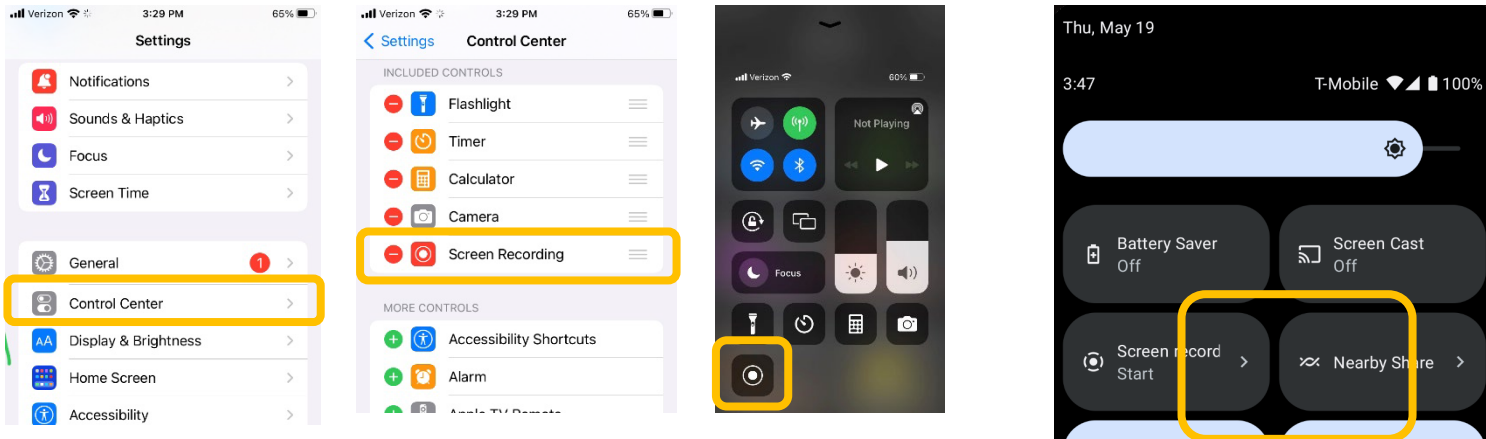


Green Banner when in Airplane Mode

Teaching with the Apps

This is a video.
Click on it to
start it!

- Take screenshots
- Share with students/teachers or entire class, or post to social media.
- Create (narrated) videos through screen capture:



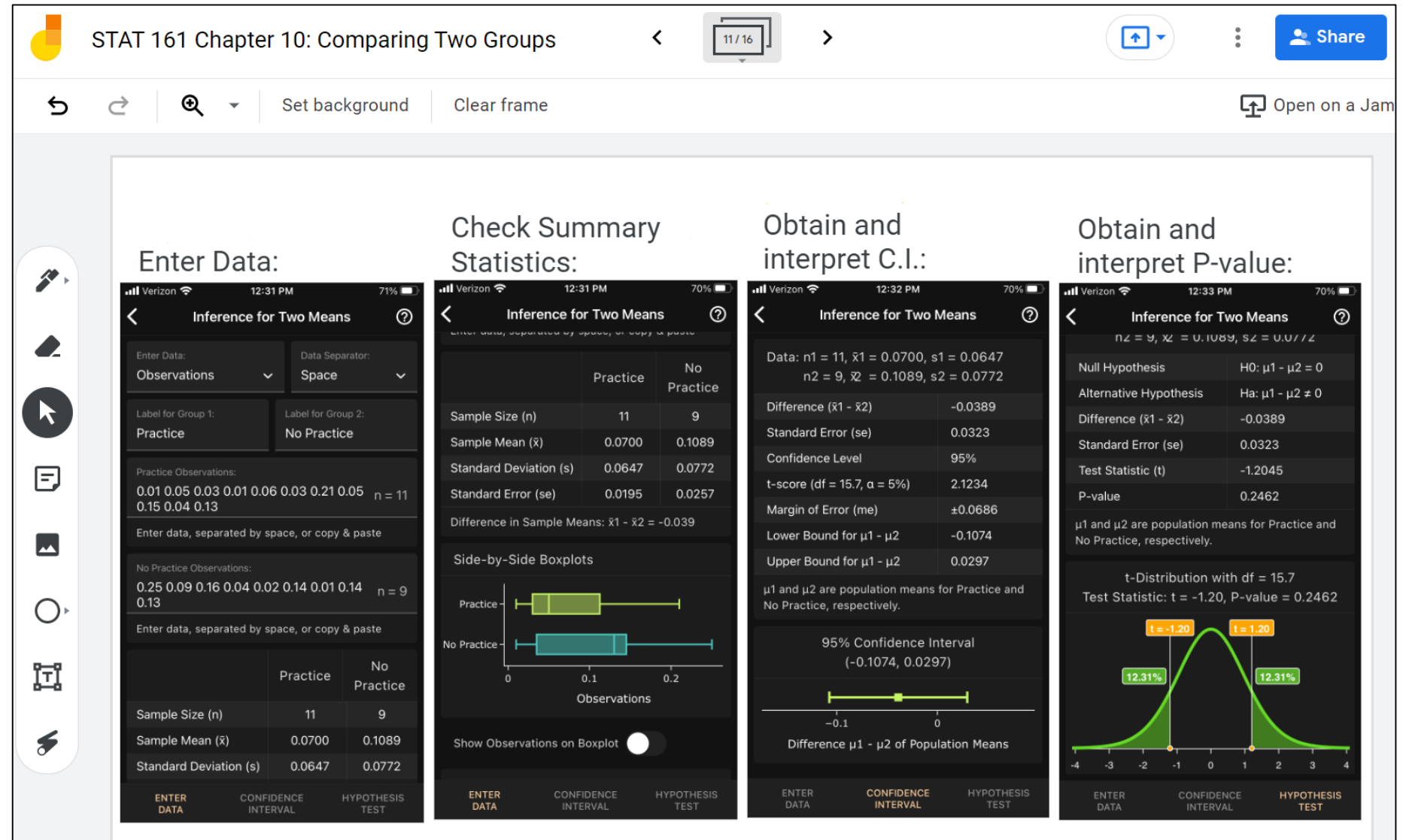
1. On iOS, go to *Settings*, then *Control Center*
2. Activate *Screen Recording*
3. To start screen recording, swipe up from the bottom

On Android, swipe down from the top twice to see if your phone supports Screen Cast

Teaching with the Apps

Include screenshots in documents.

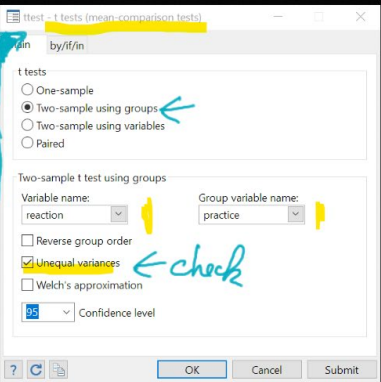
Here is a copy of my Jam-board that I share with students after class.



Teaching with the Apps

- I show output of other software (here STATA) only after having talked about the app:

STATA: You can get the C.I. via the dialog. Statistics → Summary, Tables and Tests → Classical Tests of Hypotheses → t-test OR you use the commands below:



```
. import delimited "C:\Teaching\STAT161\stopClockReaction.csv"
(2 vars, 20 obs)
. ttest reaction, by(practice) unequal
```

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
no	9	.1088889	.0257361	.0772082	.0495414	.1682364
yes	11	.07	.0194936	.0646529	.0265656	.1134344
combined	20	.0875	.0159749	.0714419	.0540642	.1209358
diff		$\bar{x}_1 - \bar{x}_2$.0388889	.0322854		- .0296657	.1074434

diff = mean(no) - mean(yes)
Ho: diff = 0
Satterthwaite's degrees of freedom = 15.683

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.8769 Pr(|T| > |t|) = 0.2462 Pr(T > t) = 0.1231

Note: STATA put the no practice group first, but results are the same

Output from commercial software such as Minitab, STATA, SPSS, JMP, but also from R is easier to understand and process once students have seen the app.

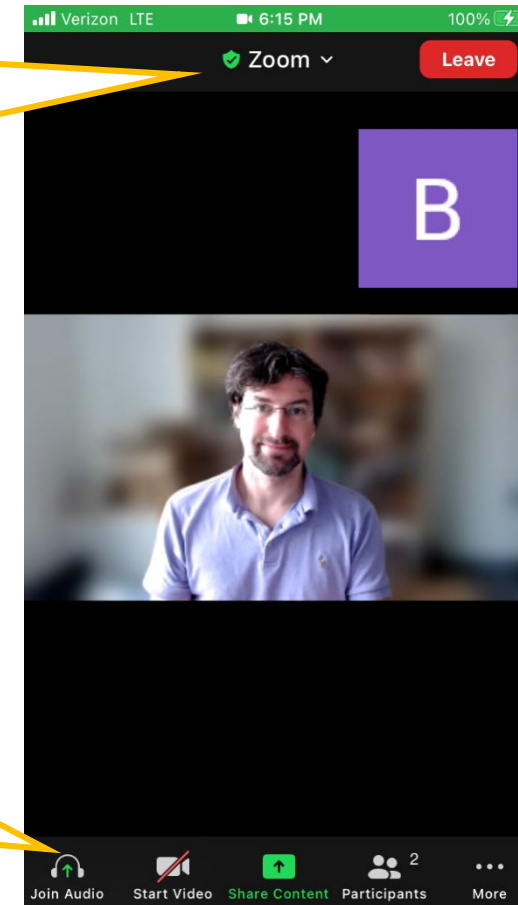
Teaching with the Apps

- Projecting the app on screen (either in class or remotely)

1. Start a Zoom call from your computer
2. Join the Zoom call on your phone (with no audio and no video)
3. On computer, “admit” phone and make it a “Co-Host”

Joined Zoom meeting from my phone.

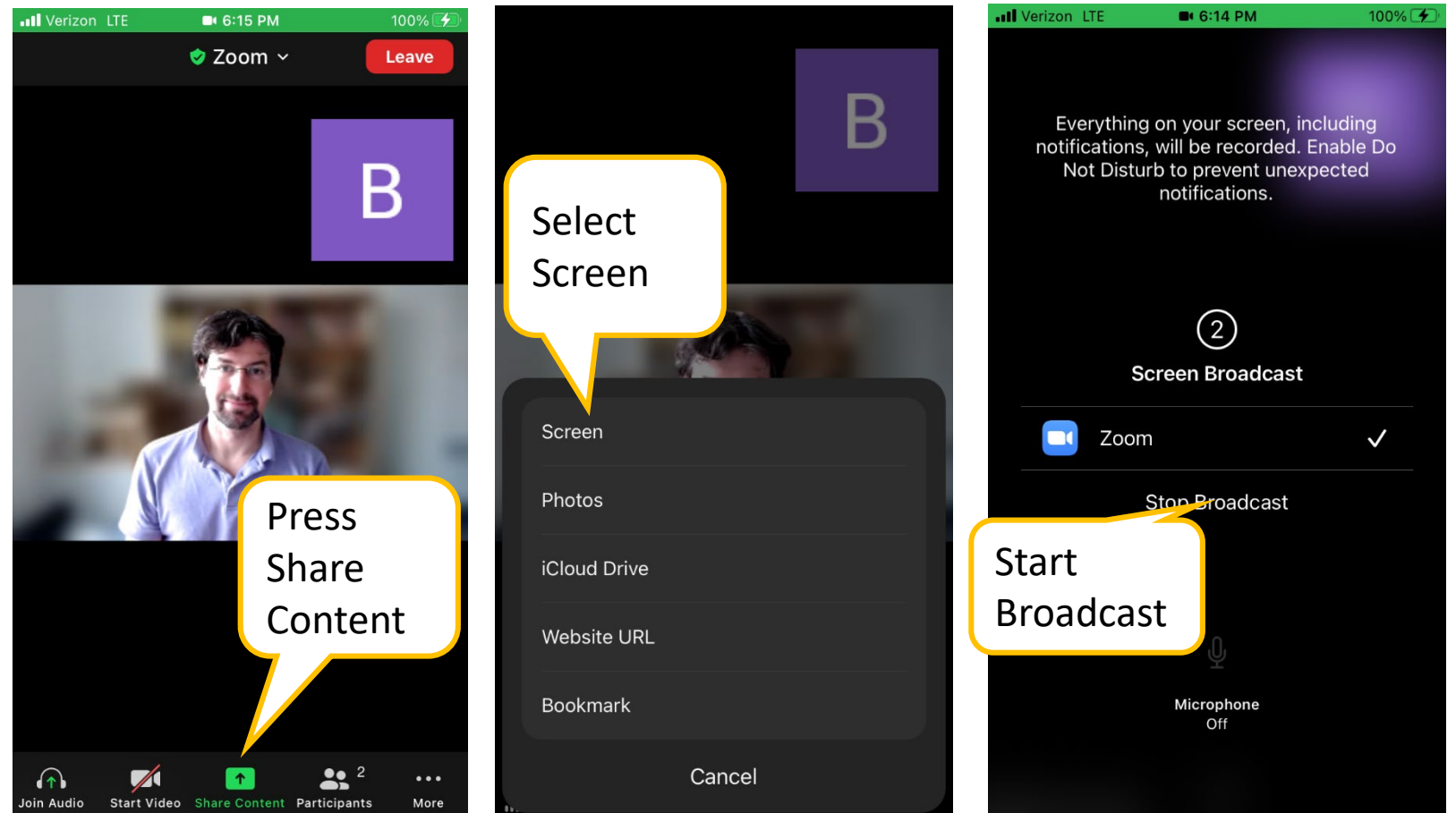
Select no audio and no video



Teaching with the Apps

- Projecting the app on screen (either in class or remotely)

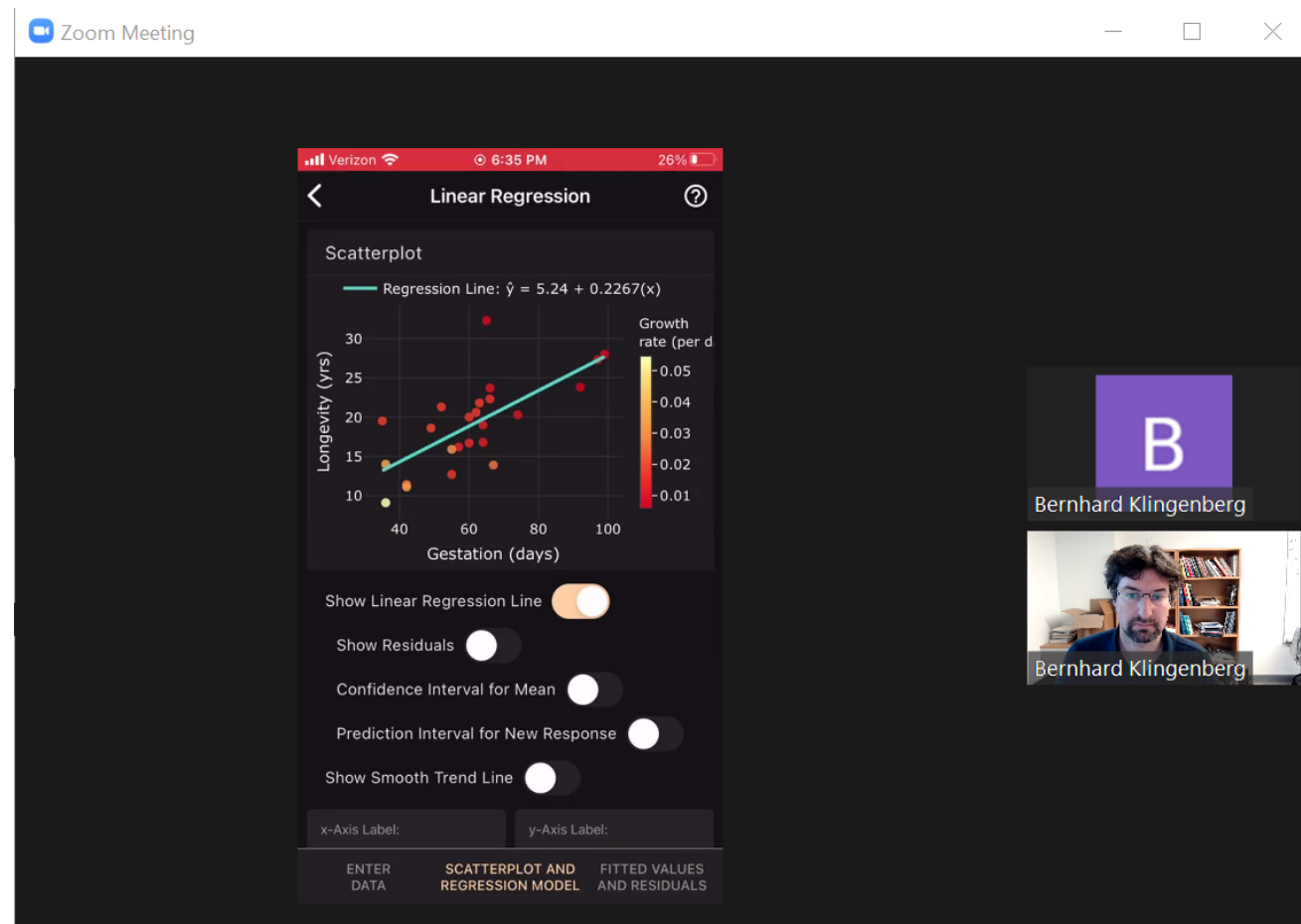
4. Press Share Content on your Phone
5. Select “Screen”
6. Go to the app to start broadcasting it



Teaching with the Apps

- Projecting the app on screen (either in class or remotely)

7. On your phone, go to the content you want to share (e.g., the Art of Stat App) Select “Screen”
8. Present on your phone



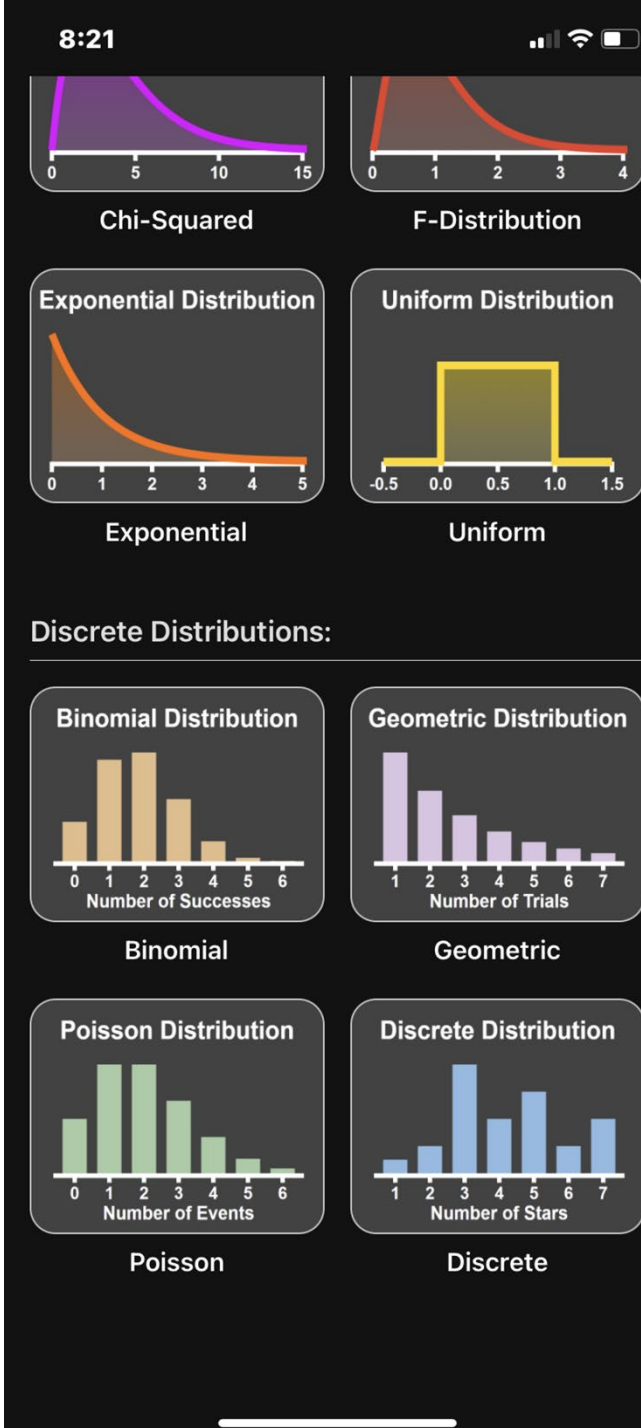
This is a video
demoing the
Linear
Regression
app and
multivariate
thinking.

Click on it to
start it!

Teaching with the Apps

- Students might get carried away a bit:

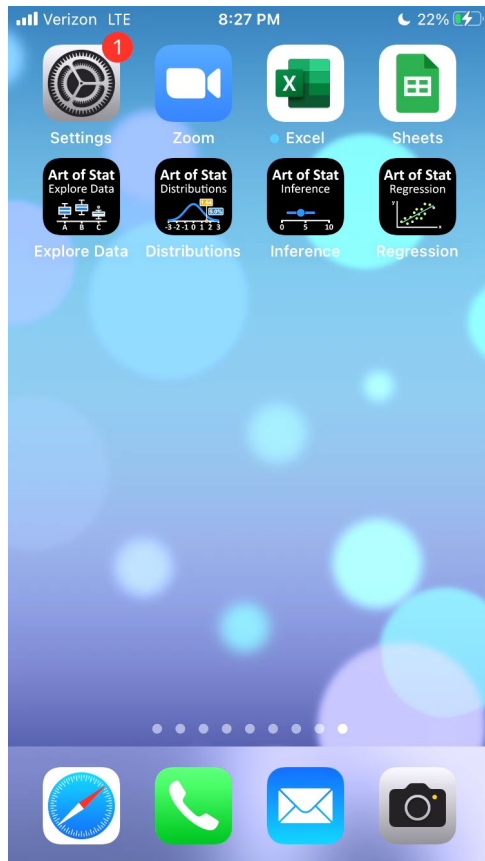
This is a video.
Click on it to
start it!



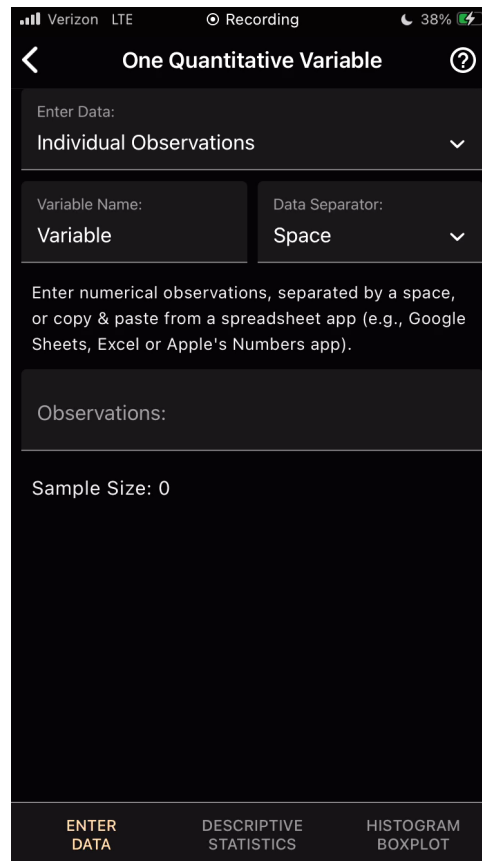
Teaching with the Apps

These are short videos to show how to upload data into the app. Click each on to play.

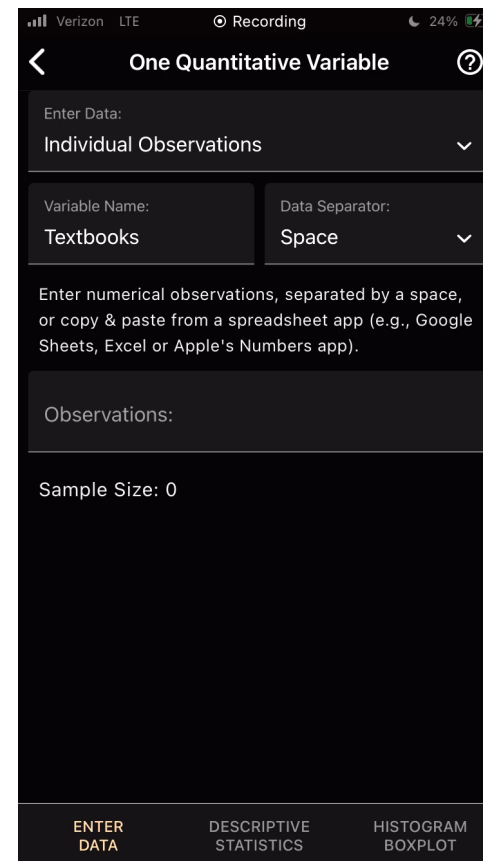
■ Data Entry



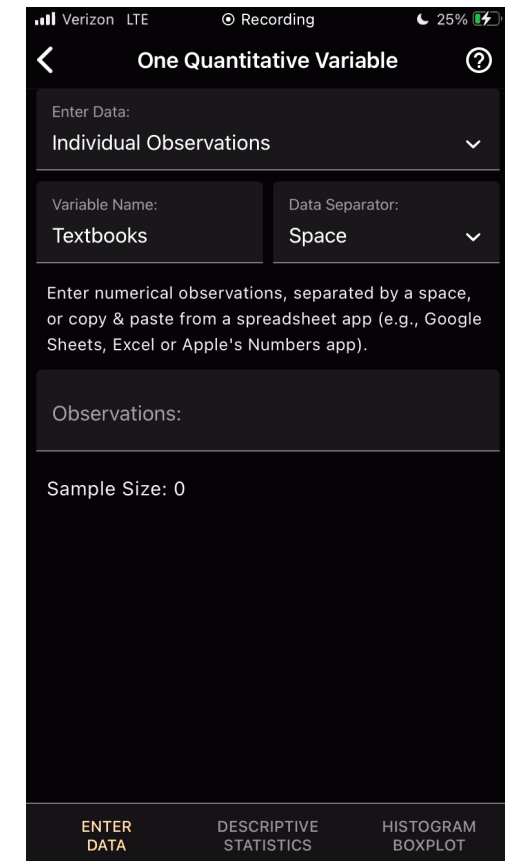
Type in data



Copy & Paste

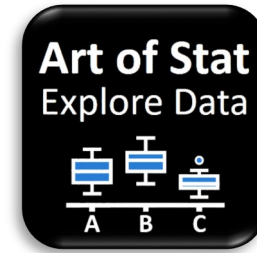


Upload CSV File



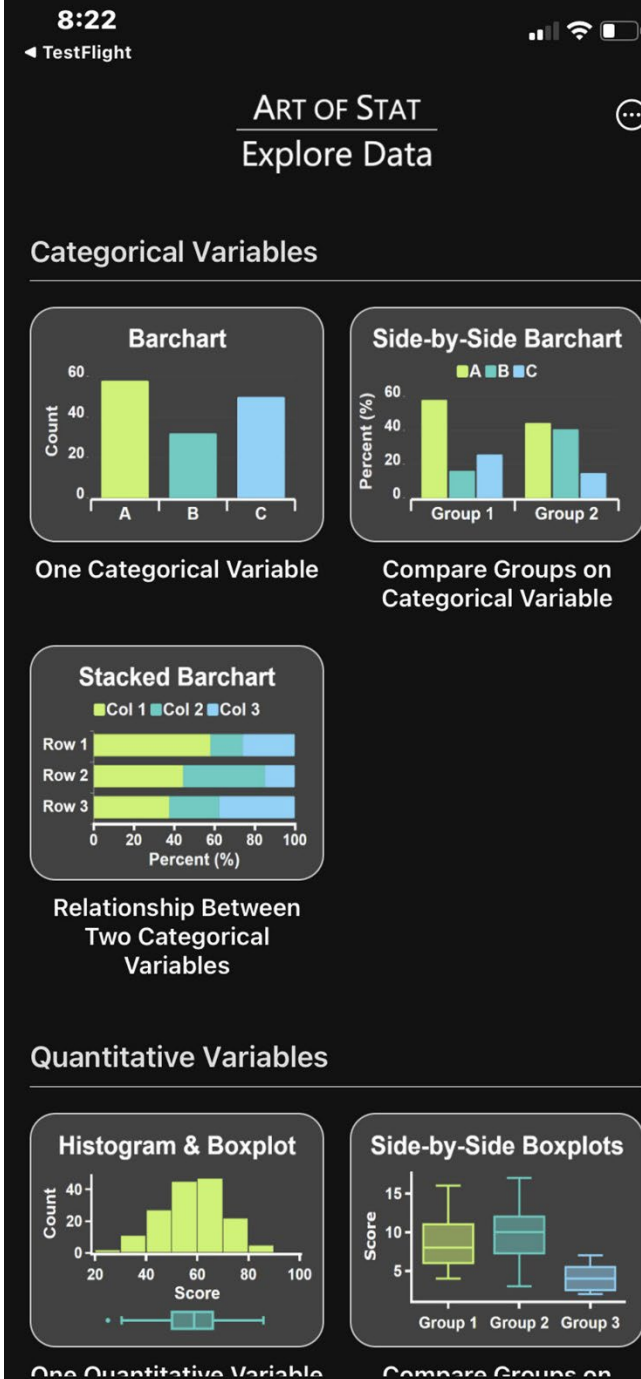
Sample Dataset

Explore Data



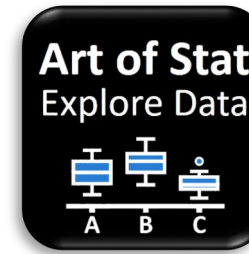
- Pre-implemented Dataset: Online Lending
 - Data on 200 randomly sampled loans made through a website

Loan	Loan Amount	Duration	Interest Rate	Grade	Home	Income	FICO Score
1	35000	36	13.6	C	rent	160000	710
2	12000	60	16.1	C	mortgage	65000	675
3	15000	36	11.8	B	rent	34000	690
4	3025	36	15.1	C	mortgage	150000	740
5	9000	36	14.5	C	own	70000	705
6	20000	36	6.5	A	mortgage	100000	785
7	12500	60	7.8	A	rent	41600	700
8	5000	36	18.9	D	mortgage	44000	685
9	21000	60	16	C	mortgage	80000	685
10	35000	36	14.5	C	mortgage	90000	745



Explore Data

- Analyze the variable “Grade of the Loan”



Verizon LTE 12:17 PM 63%

One Categorical Variable

Enter Data:
Counts

Variable Name:
Variable

Number of Categories:
3

Provide Labels for Categories:
Category 1: Cat 1
Category 2: Cat 2
Category 3: Cat 3

Enter Data

Counts

Open CSV File

Individual Observations

Open Sample Datasets

Clear Cell Counts

Reset to Defaults

Cancel

Press the dropdown for “Enter Data”

Select Open Sample Datasets, then Online Lending

Verizon LTE 12:18 PM 63%

One Categorical Variable

Enter Data:
Open Sample Datasets

Select Dataset:
Online Lending

Loan	Loan Amount	Duration	Interest Rate
1	35000	36	13.6
2	12000	60	16.1
3	15000	36	11.8
4	3025	36	15.1
5	9000	36	14.5
6	20000	36	6.5
7	12500	60	7.8

Showing first 100 rows of the 200 rows in the CSV file, and all 8 columns.

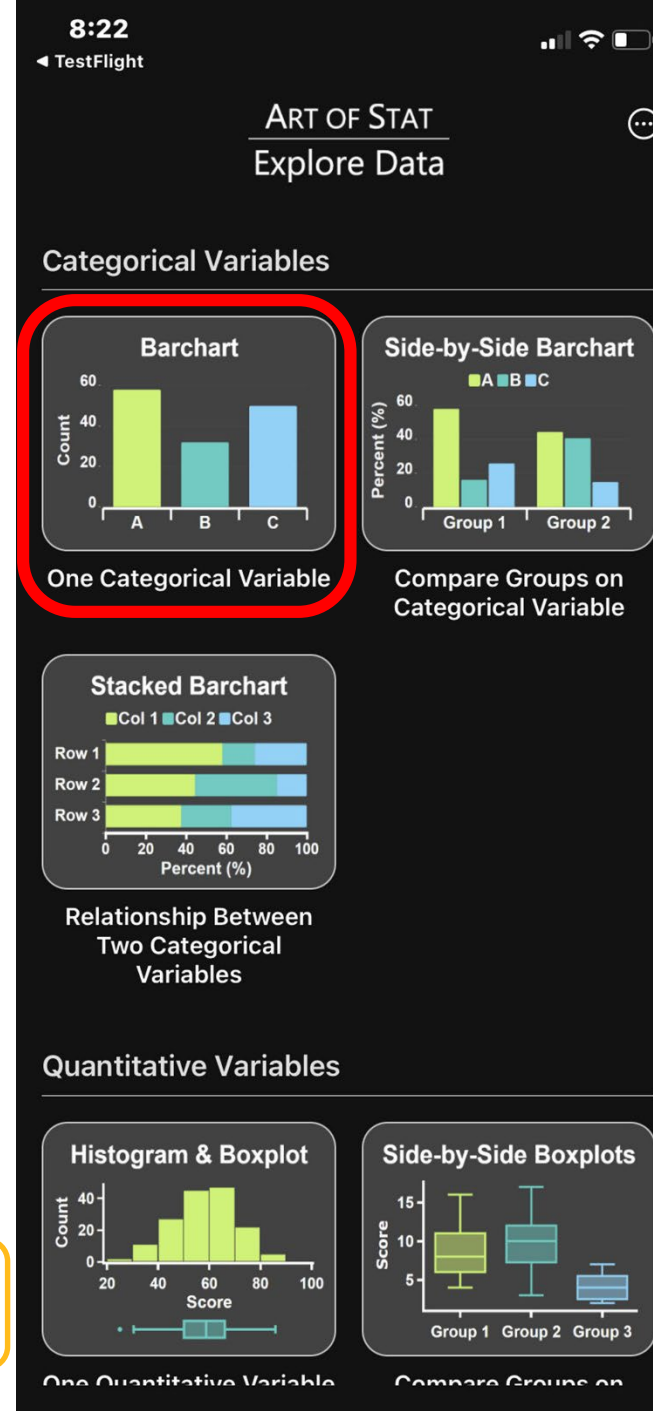
Select Variable:
Grade

ENTER DATA FREQUENCY TABLE BAR & PIE CHARTS

Selected “Online Lending”

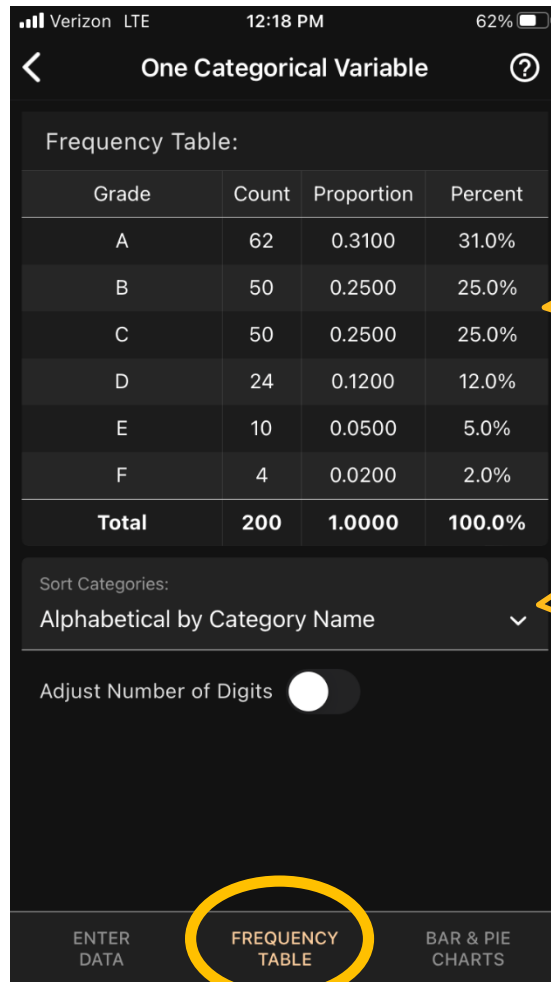
Online Lending Dataset is displayed

Select “Grade” as the Variable



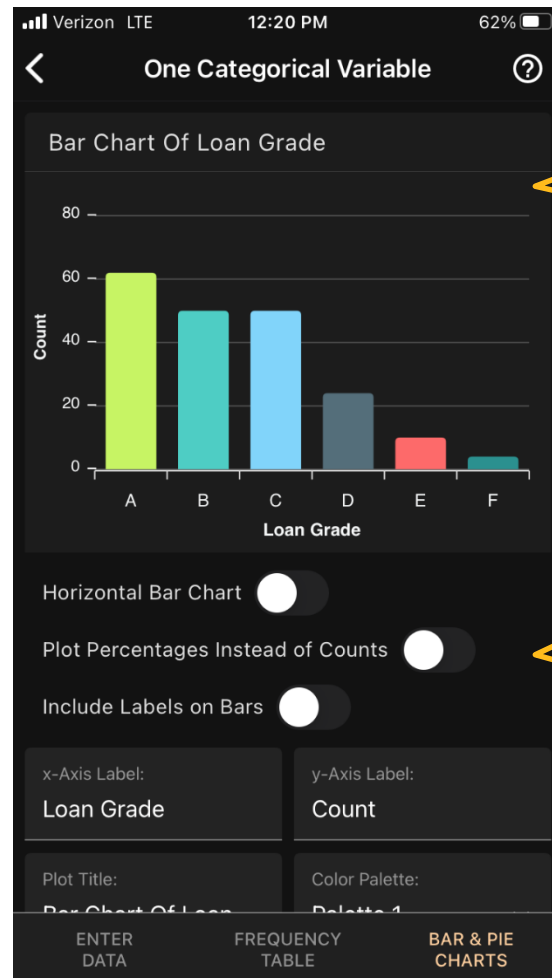
Explore Data

- Analyze the variable “Grade of the Loan”



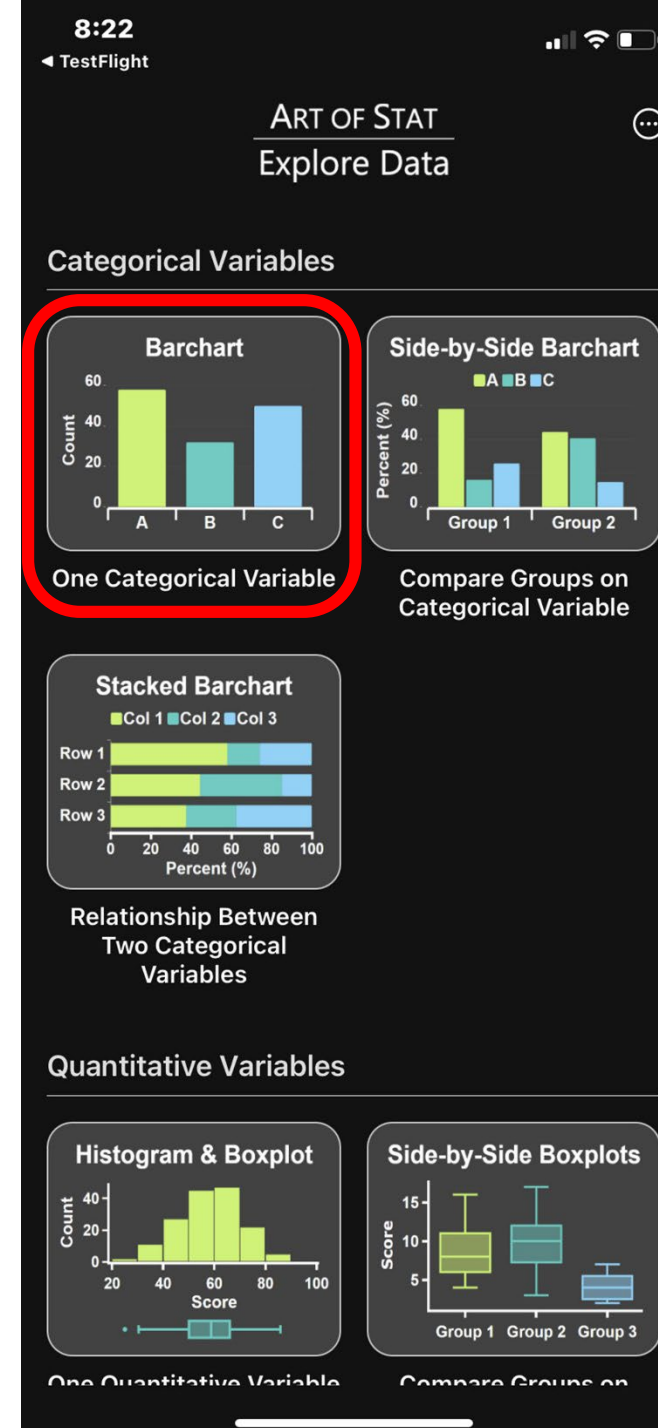
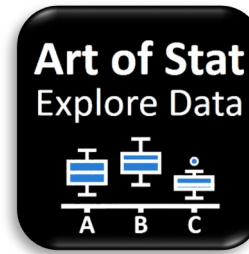
Frequency Table

Sort Grade Labels Alphabetically



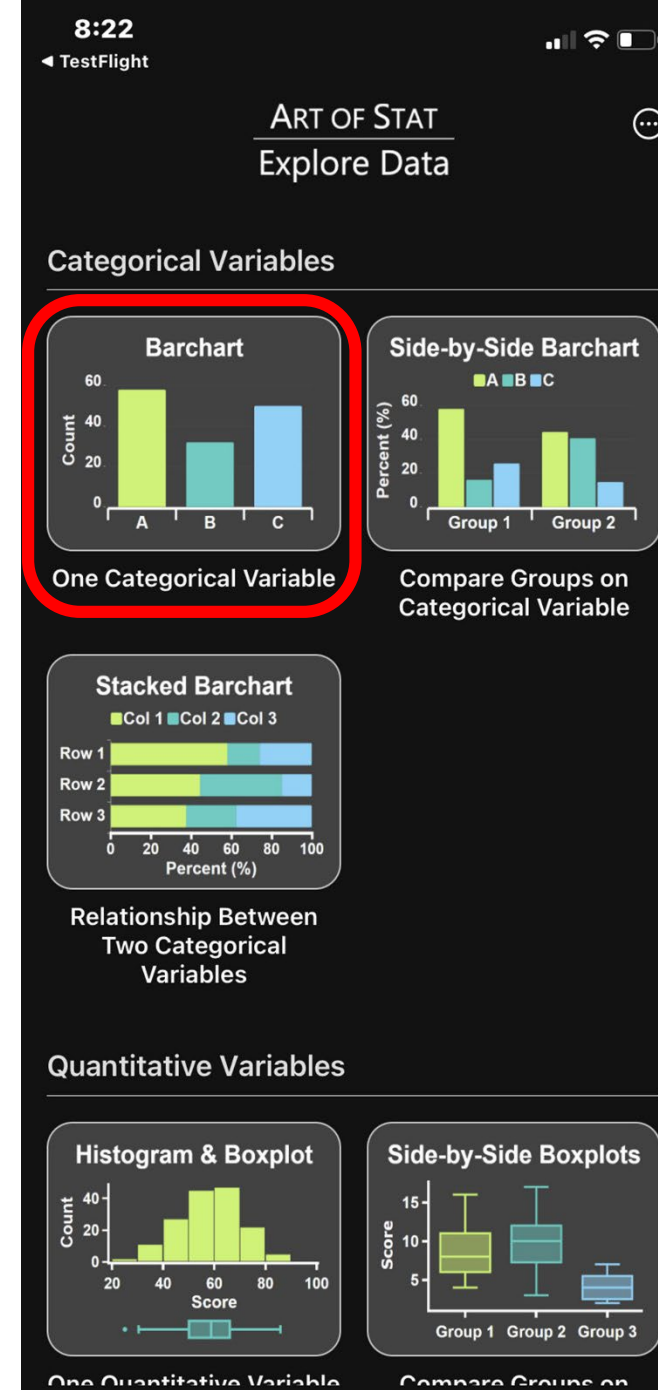
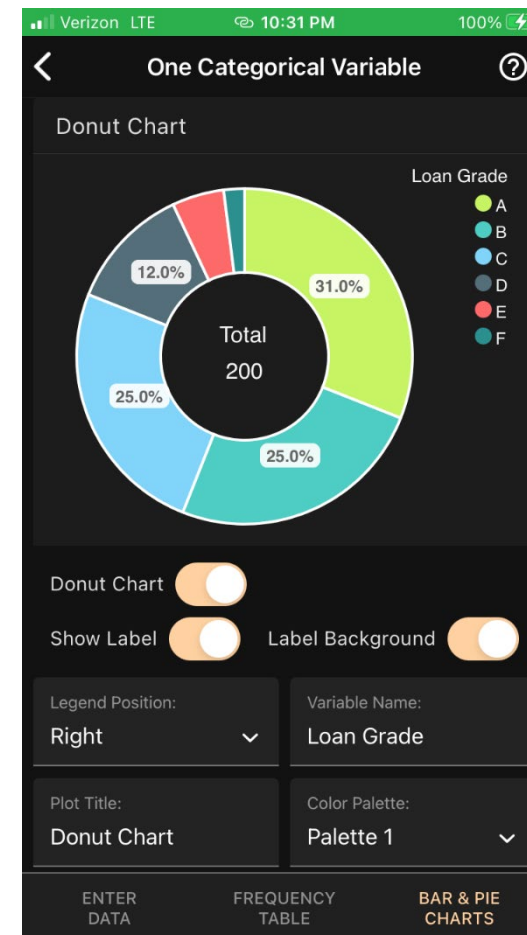
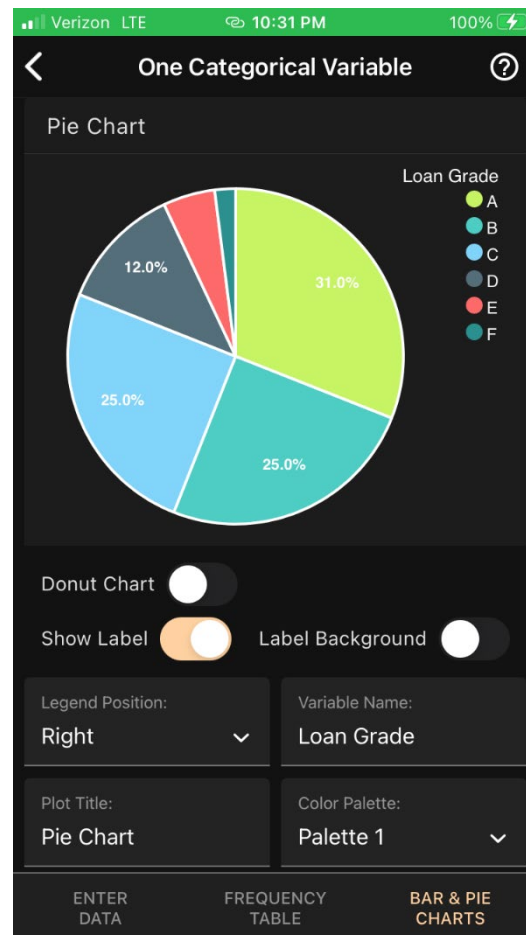
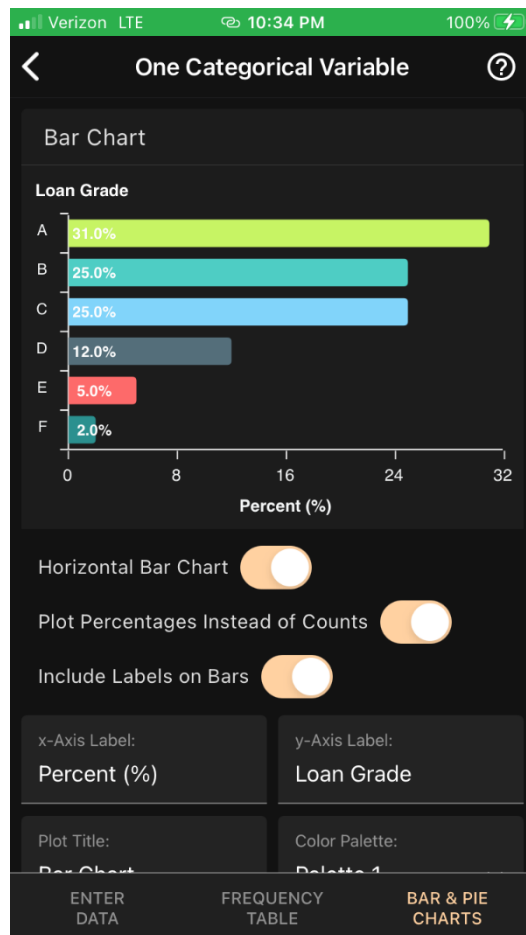
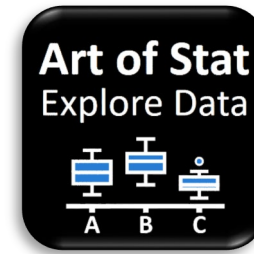
Bar Chart

Various Options



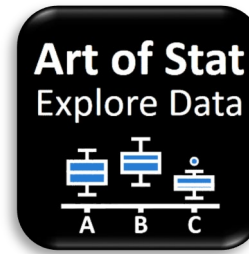
Explore Data

Other Options:



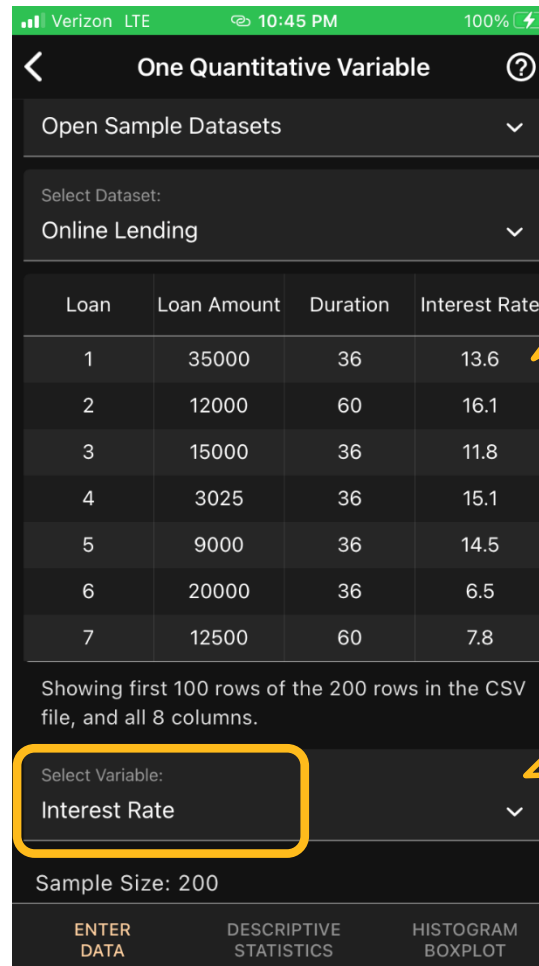
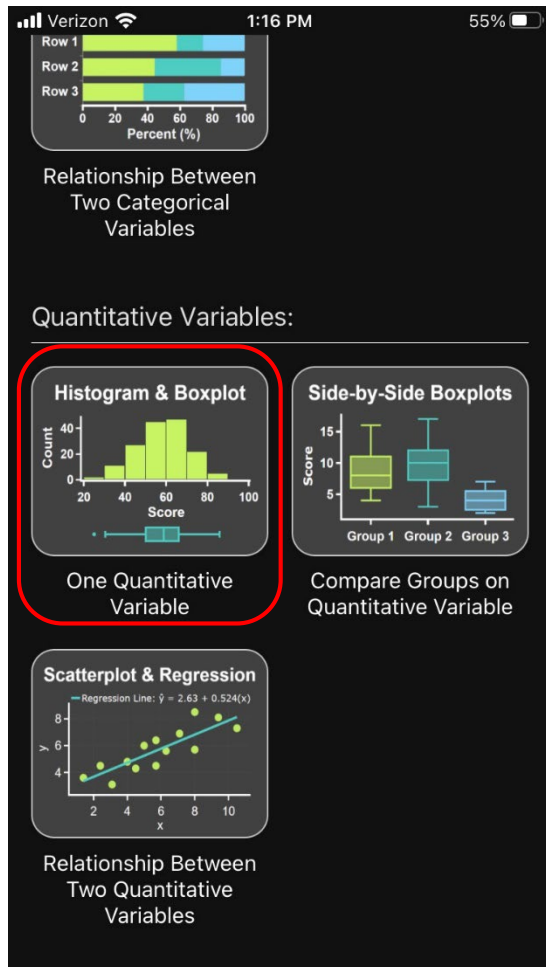
Explore Data

- Now, analyze the variable “Interest Rate”



Interest Rate

13.6
16.1
11.8
15.1
14.5
6.5
7.8
18.9
16
14.5



One Quantitative Variable

Open Sample Datasets

Select Dataset:
Online Lending

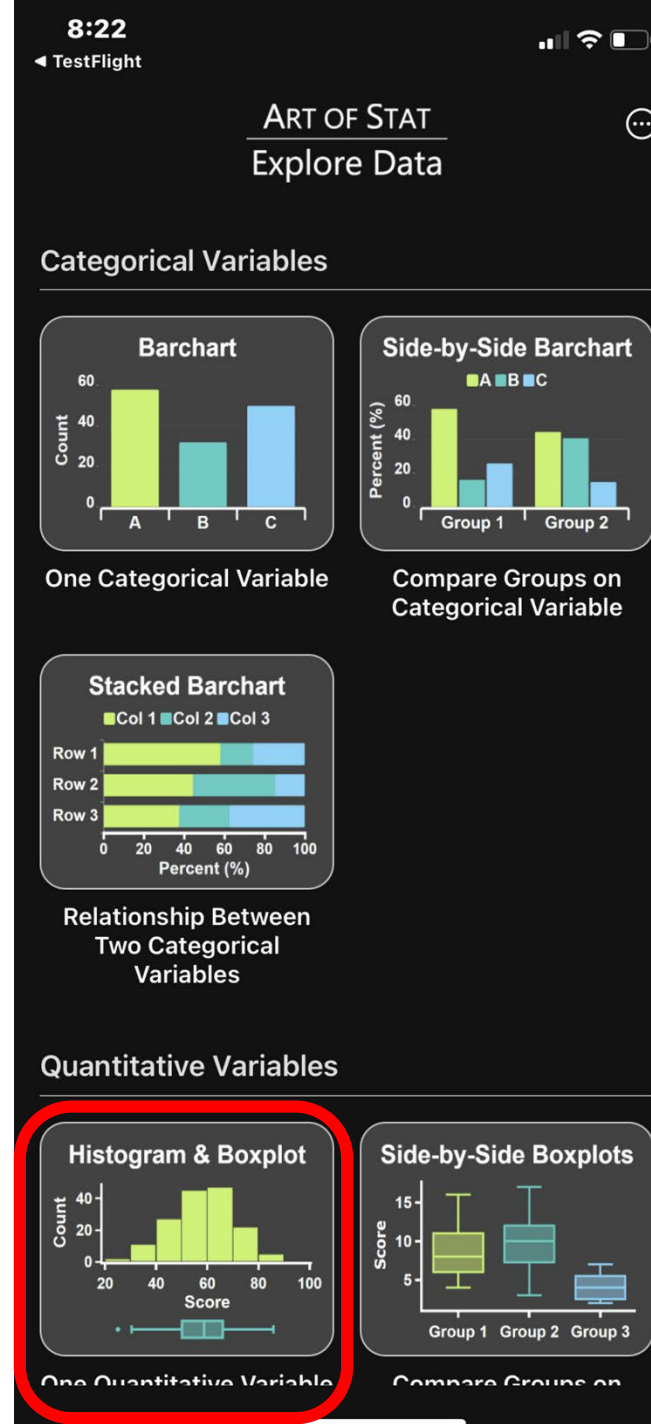
Loan	Loan Amount	Duration	Interest Rate
1	35000	36	13.6
2	12000	60	16.1
3	15000	36	11.8
4	3025	36	15.1
5	9000	36	14.5
6	20000	36	6.5
7	12500	60	7.8

Showing first 100 rows of the 200 rows in the CSV file, and all 8 columns.

Select Variable:
Interest Rate

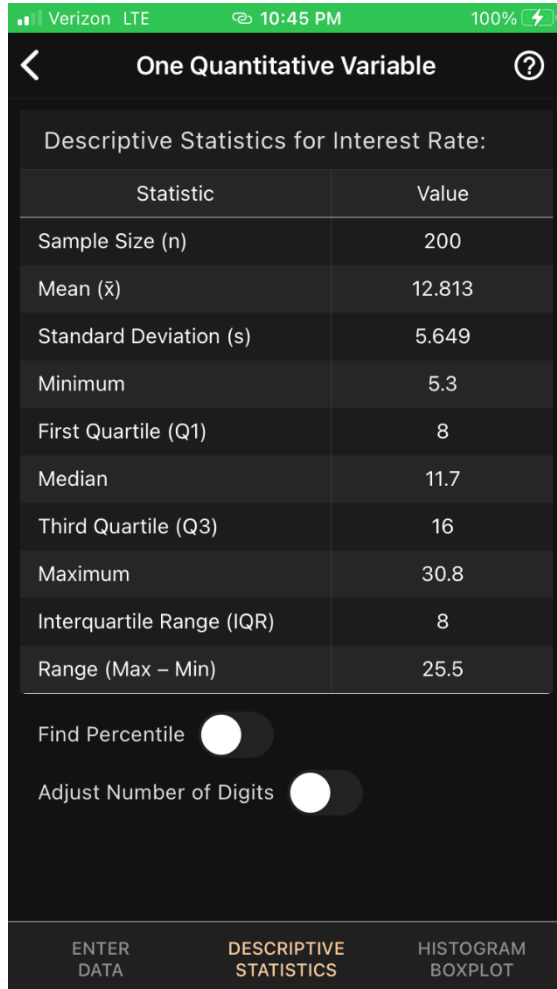
Sample Size: 200

ENTER DATA DESCRIPTIVE STATISTICS HISTOGRAM BOXPLOT

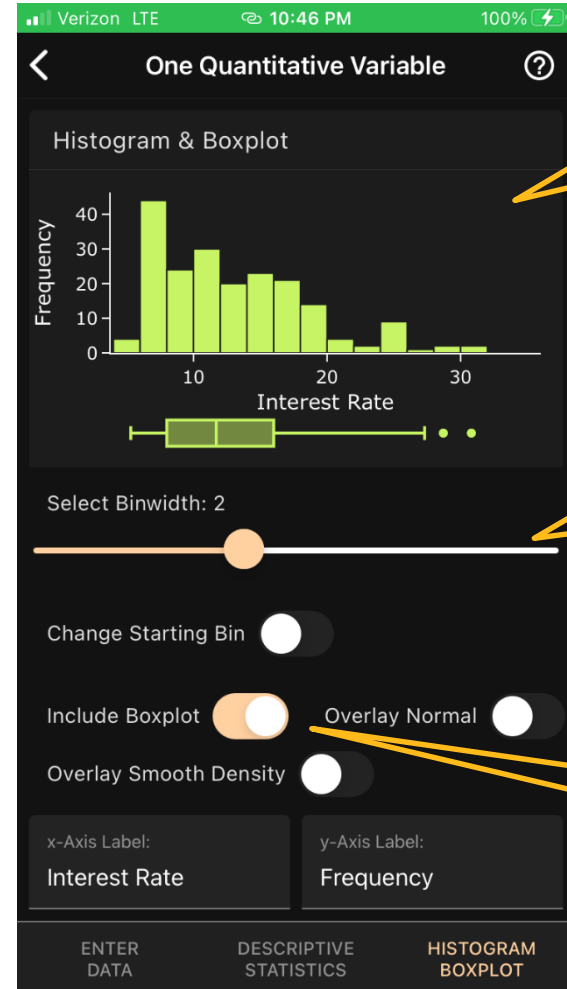


Explore Data

■ Quantitative Variable Analysis:



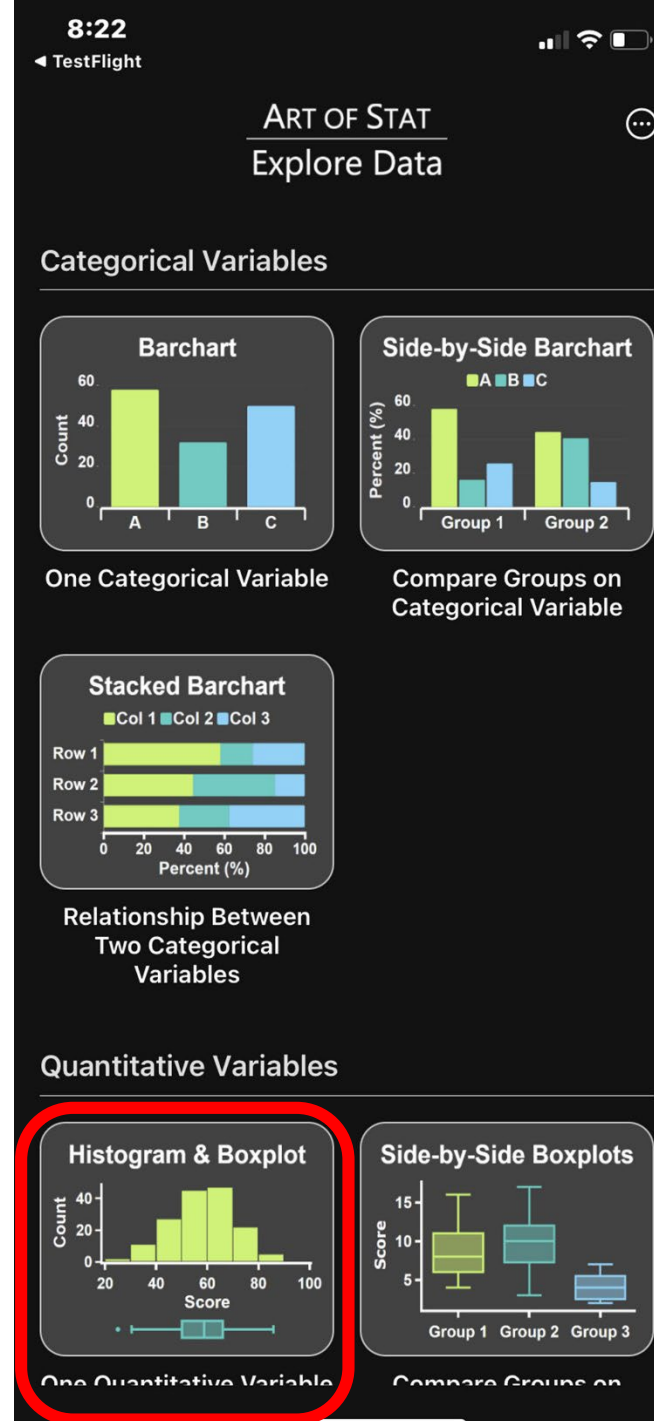
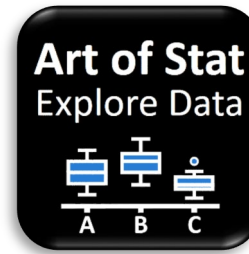
Descriptive Statistics



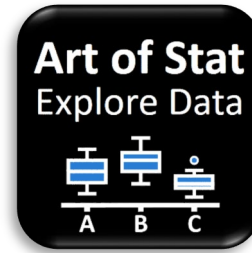
Histogram of Interest Rate

Vary bin size interactively with slider

Include Boxplot

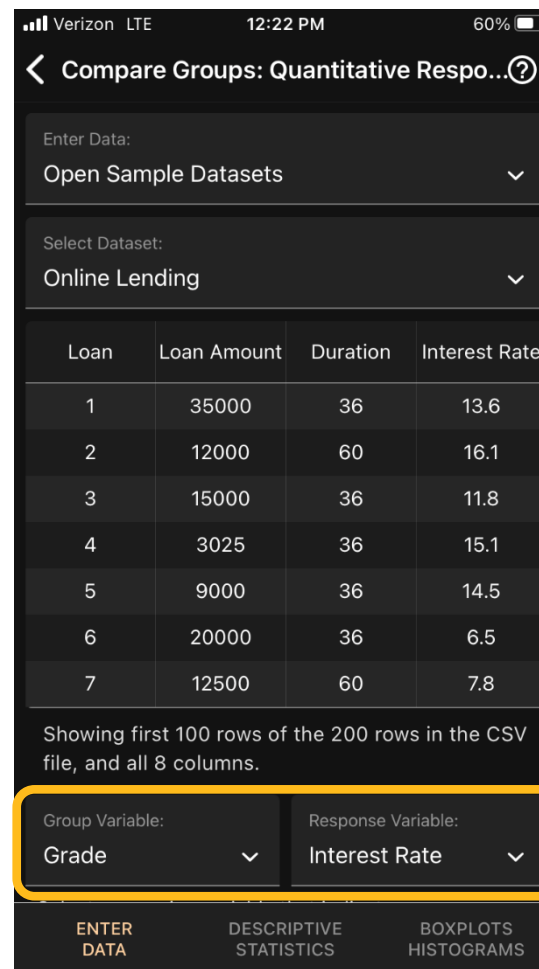
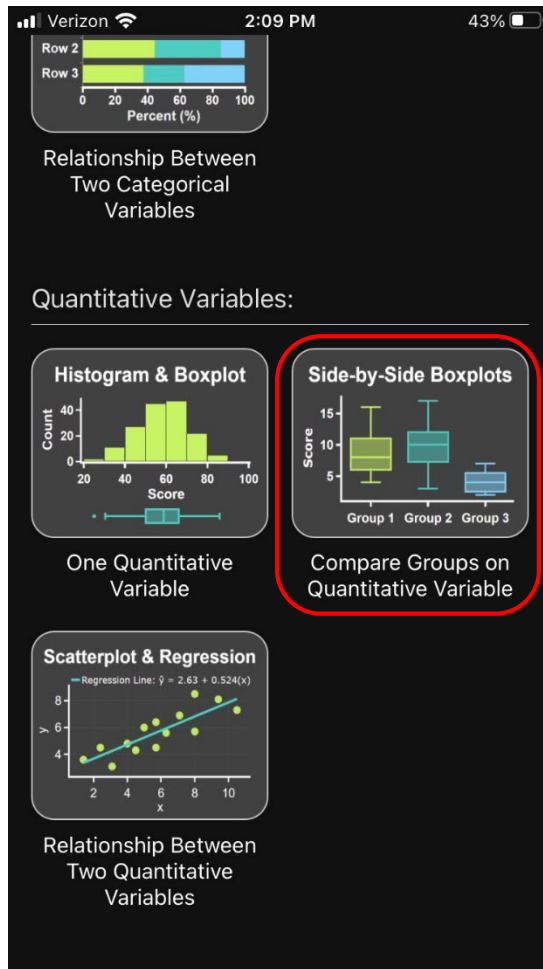


Explore Data



■ Finally: Compare Groups on Quantitative Variable

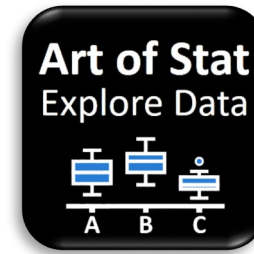
Interest Rate	Grade
13.6	C
16.1	C
11.8	B
15.1	C
14.5	C
6.5	A
7.8	A
18.9	D
16	C
14.5	C



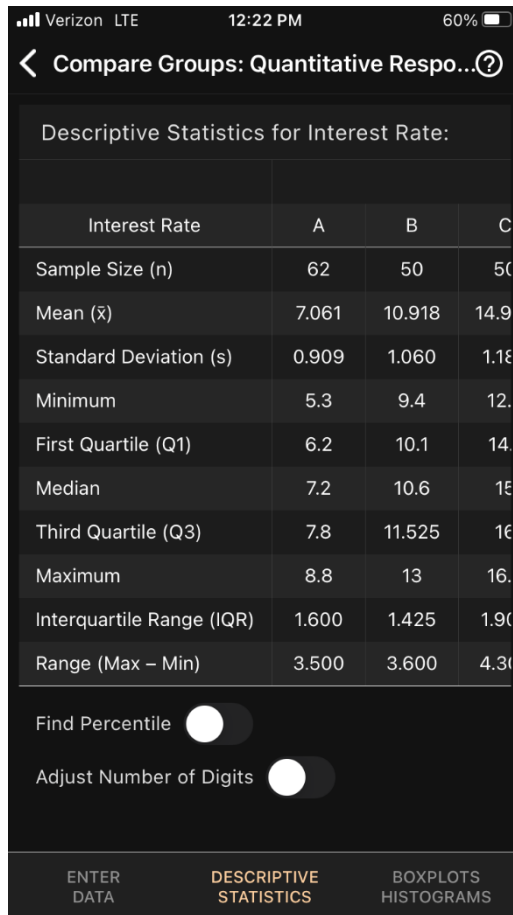
Select Grade as the Grouping Variable, and Interest Rate as the Response Variable



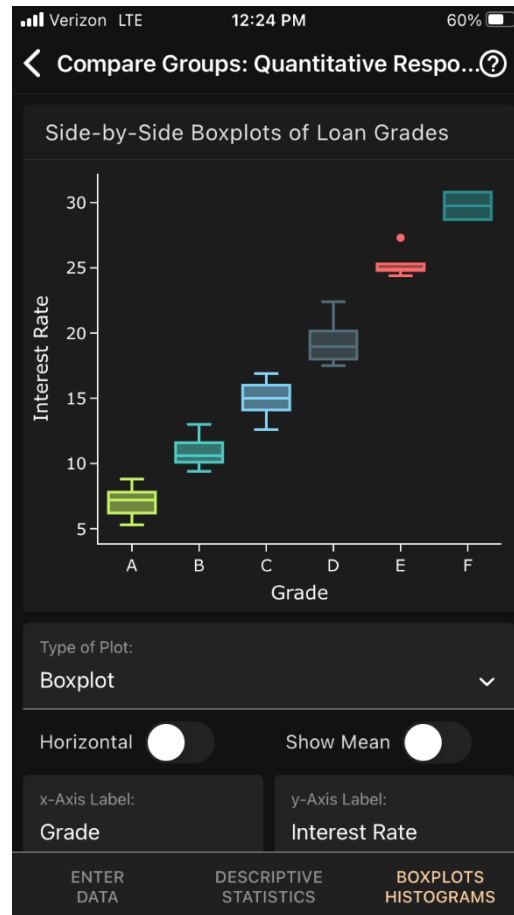
Explore Data



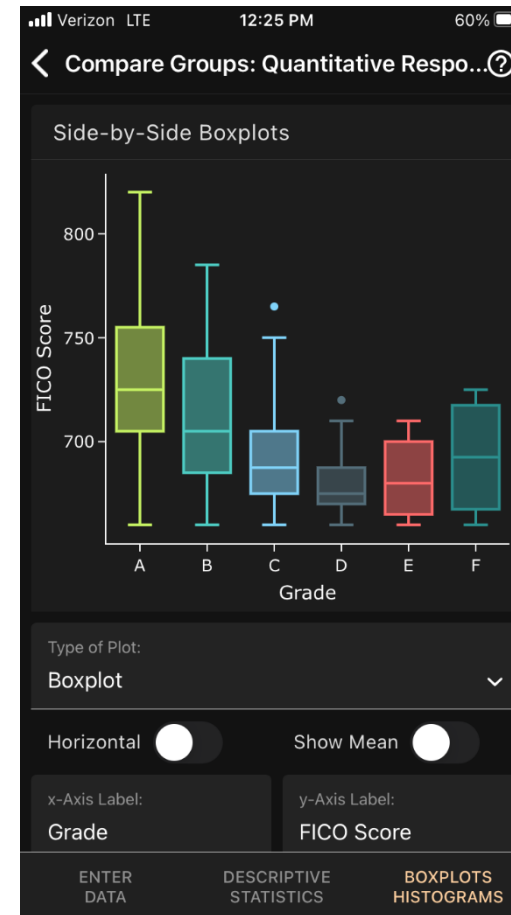
■ Finally: Compare Groups on Quantitative Variable



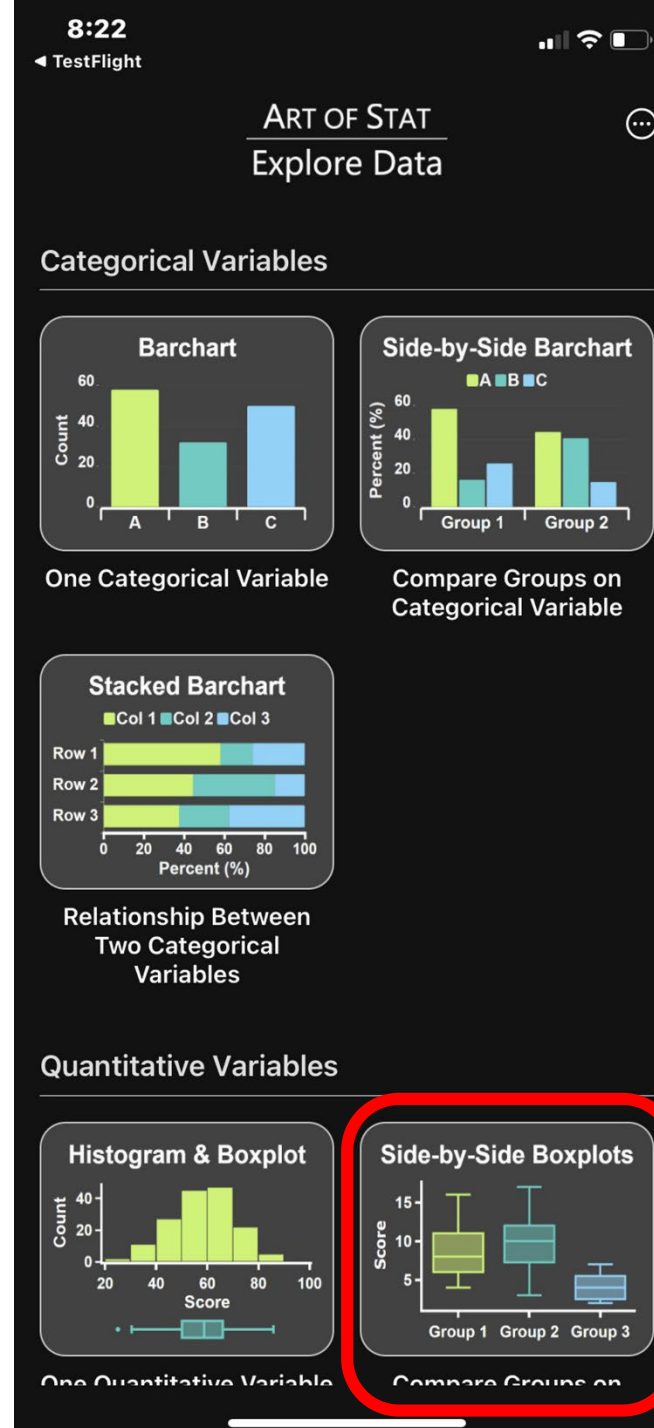
Descriptive Statistics



Side-By-Side Boxplots

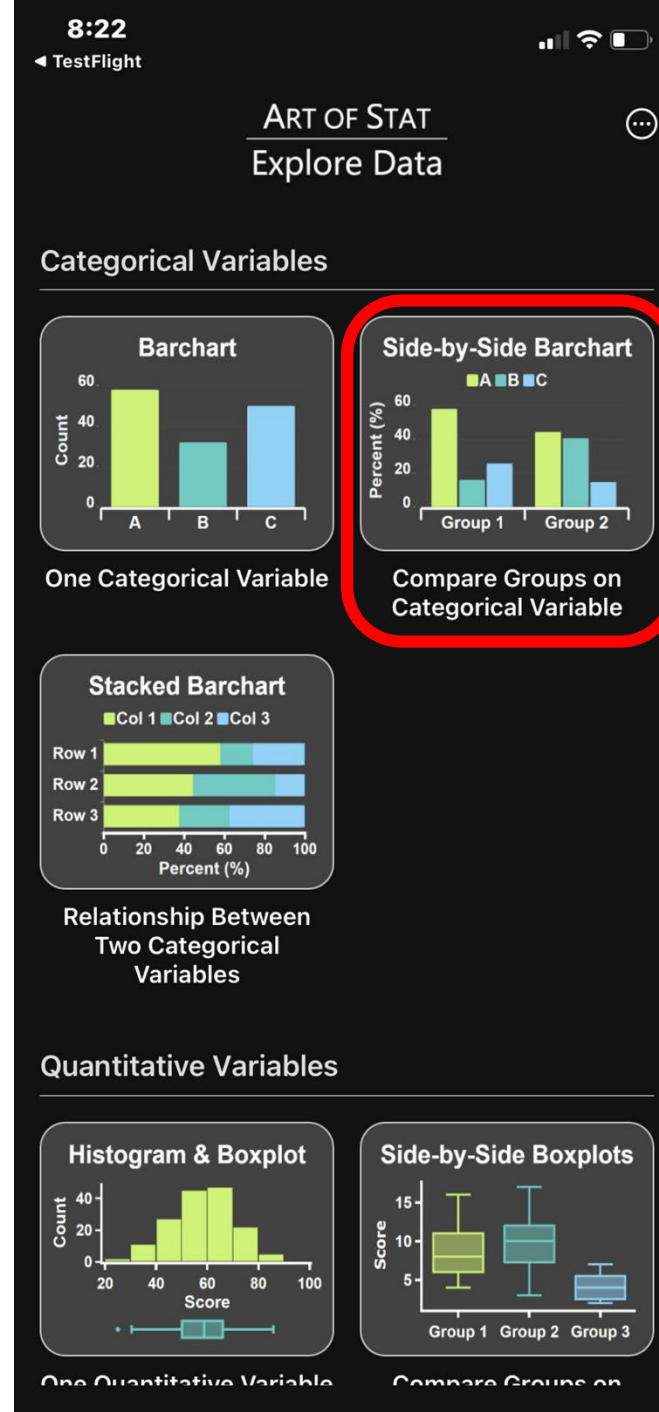
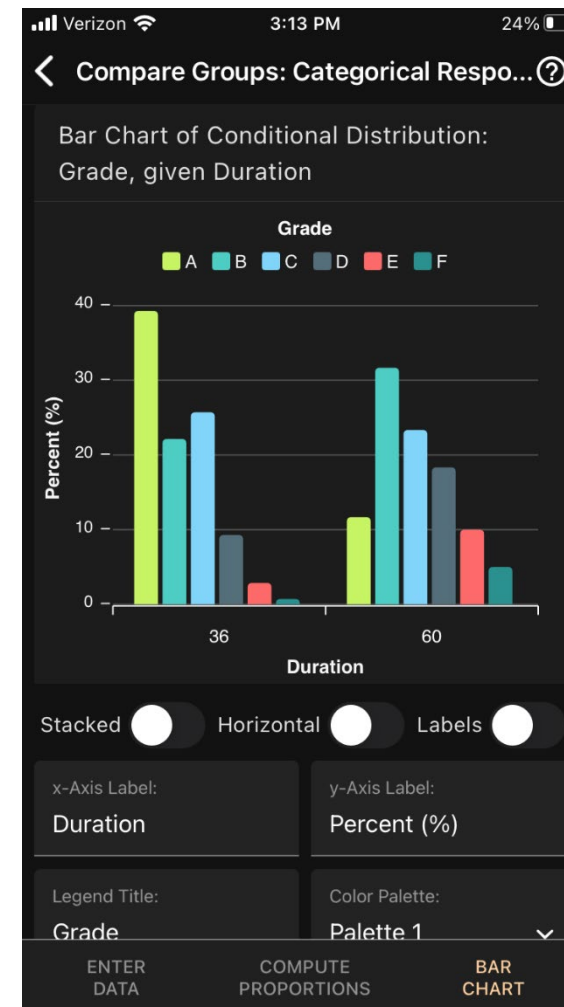
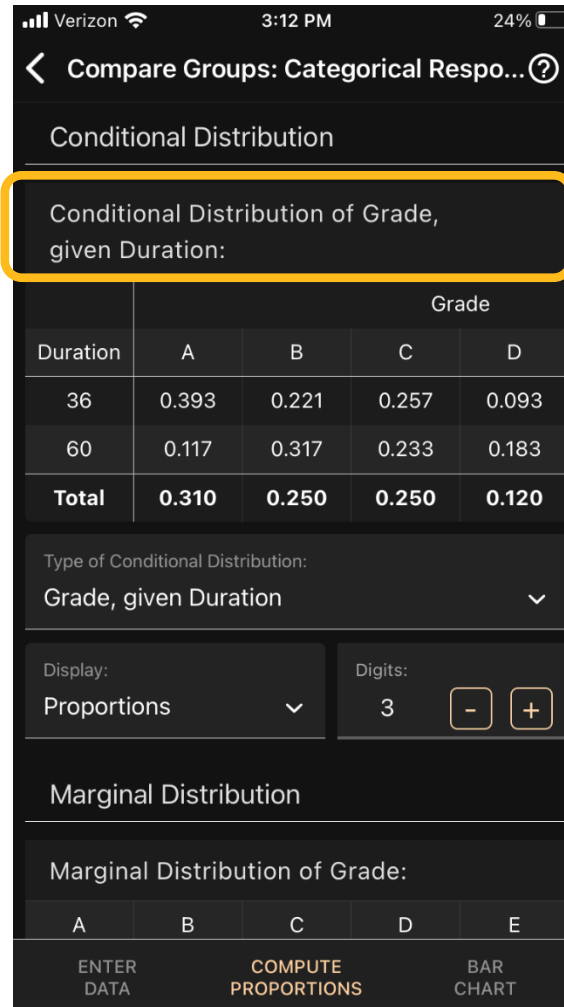
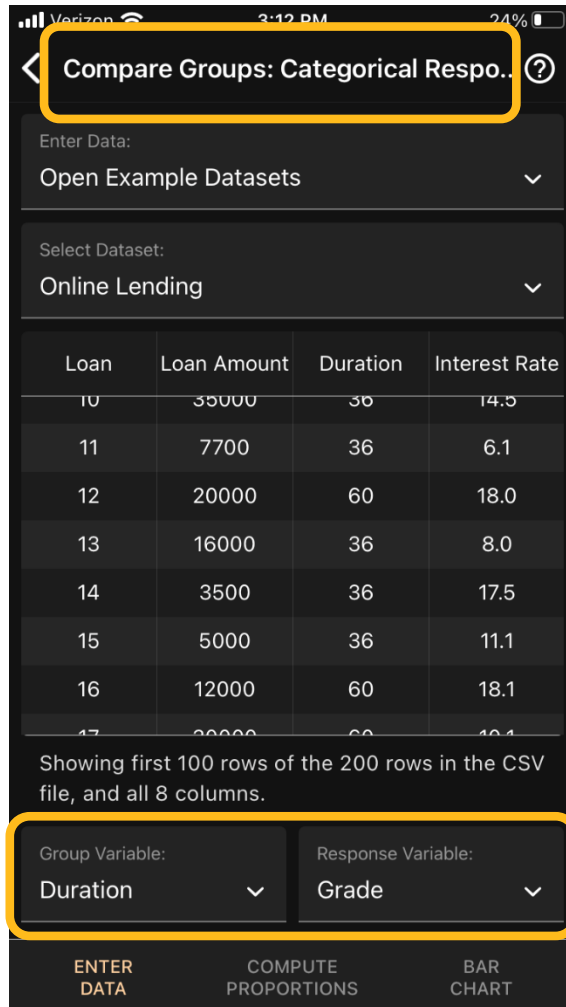
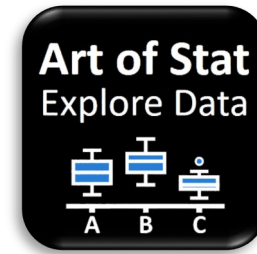


With FICO Score

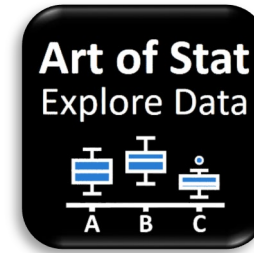


Explore Data & Inference

■ Apps handle Contingency Tables

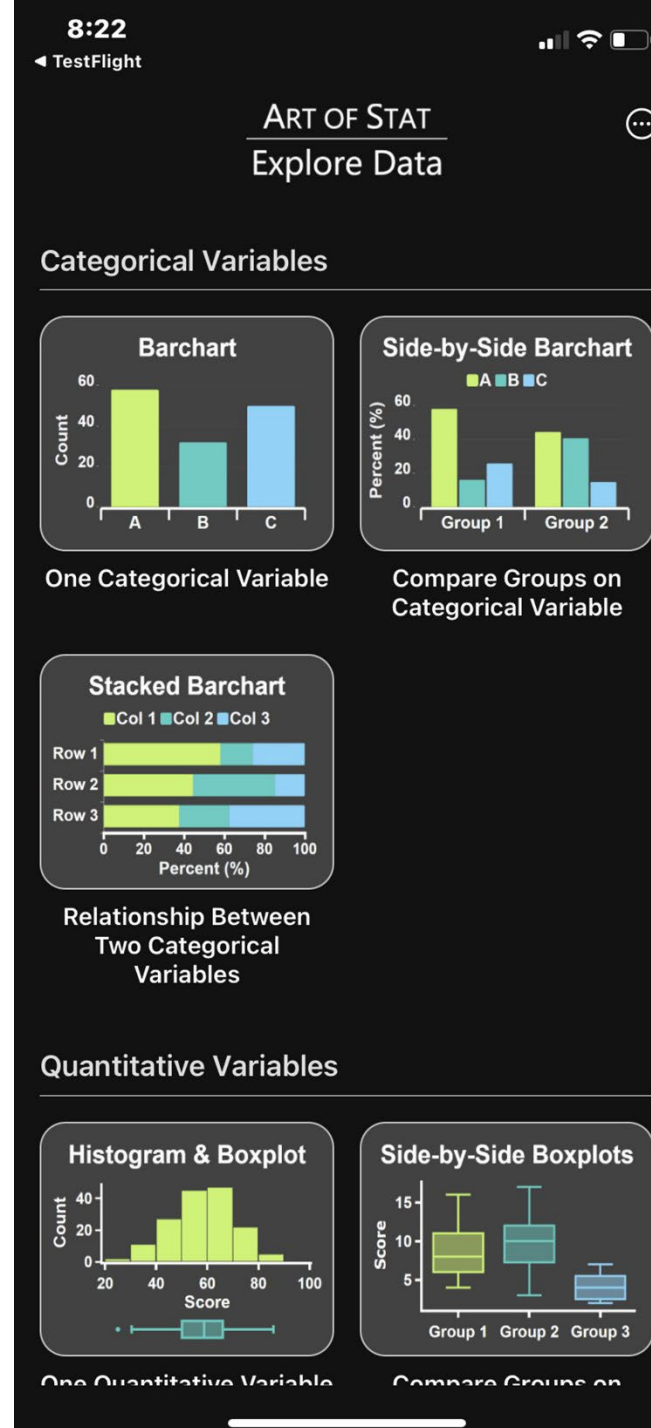


Your turn!



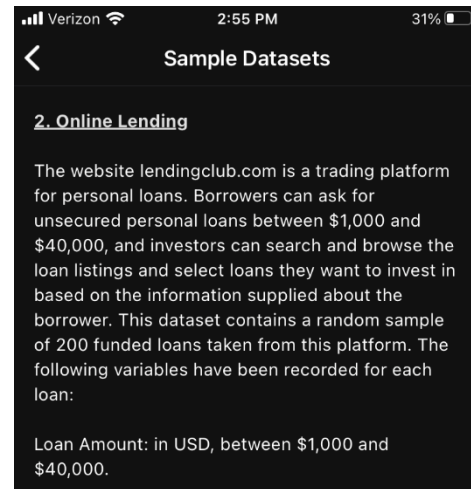
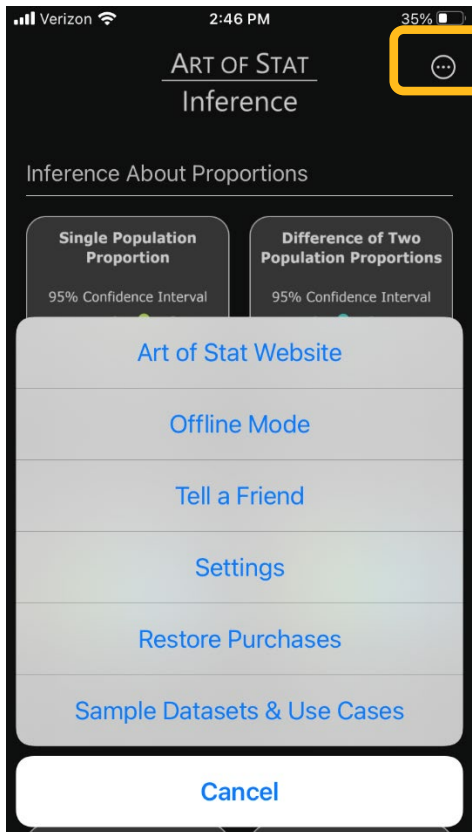
- A couple of prompts:
 - The online lending website records home ownership of loan applicants. Explore the distribution of **Home Ownership**.
 - You can either get a 36 months or a 60 months loan. Which one is more popular? Explore the distribution of **Loan Duration**.
 - Investigate the relationship btw. **Loan Duration** and **Grade**. Do 60 months loans tend to have lower grades? *
 - The online lending website records the FICO score of loan applicants. Explore the distribution of **FICO scores**.*
 - How do **FICO scores** compare across **Home Ownership**? *

* Requires unlocking

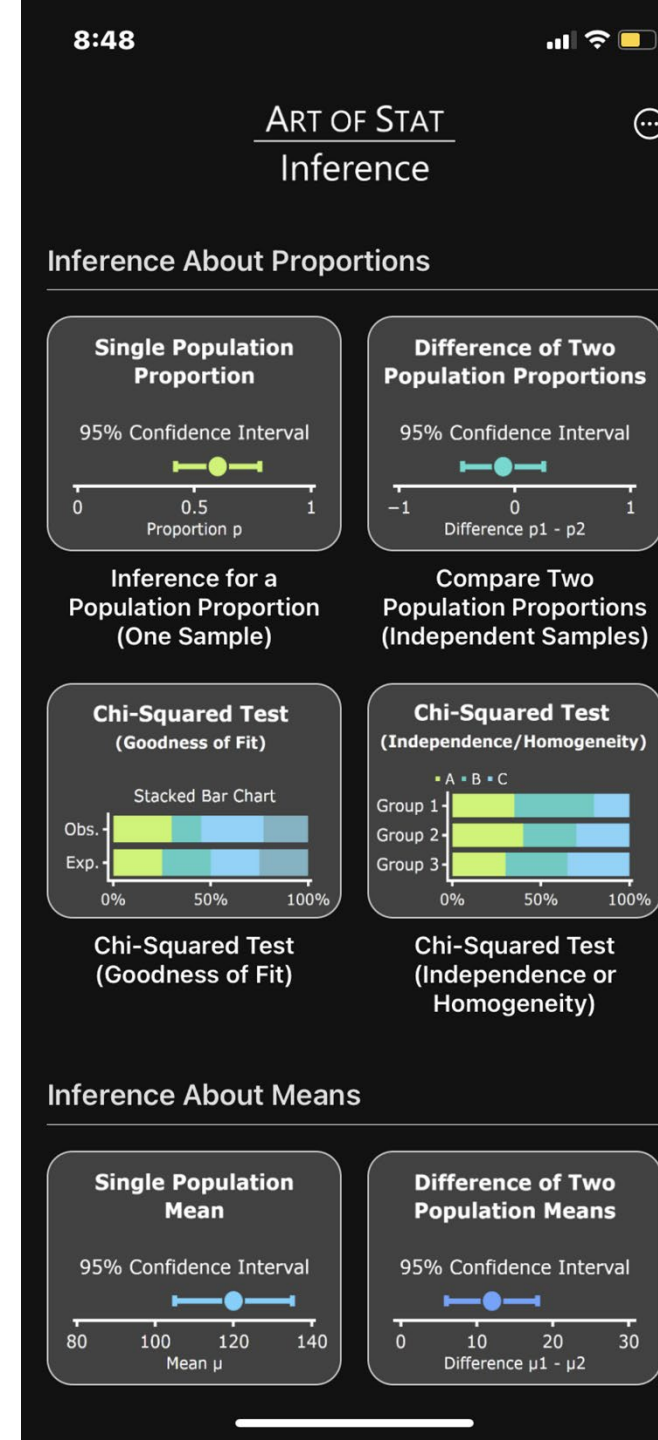
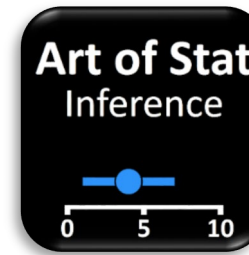
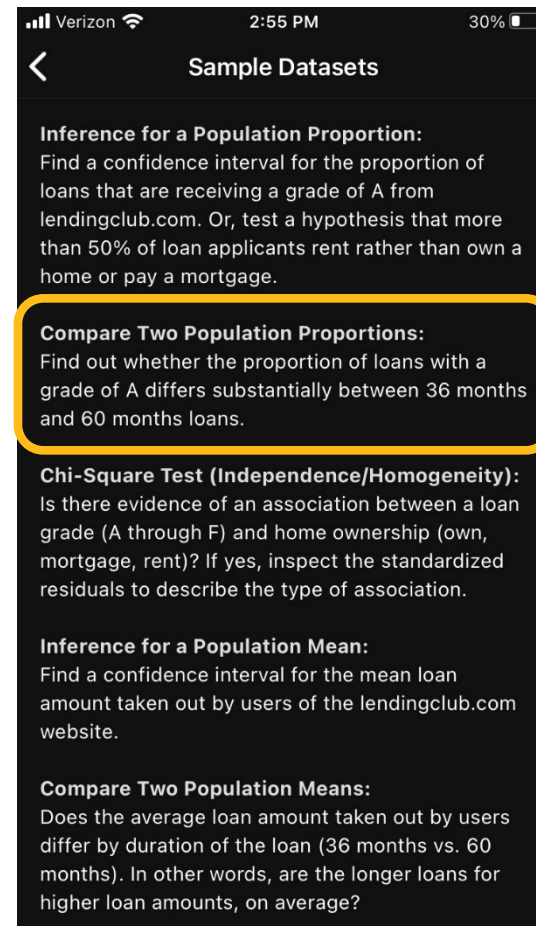


Your turn!

- The Online Lending dataset is also available as a Sample Dataset in the Inference app.



Access its descriptions and prompts for inference by going to *Sample Datasets & Use Cases*. (Press the three dots in the top right on the landing page.)



Concluding Remarks

Technology is crucial for how students and teachers interact with the material in a course:

- Concepts come to life when ...
... you can “see” and interact with them.
- Concepts and ideas become memorable when ...
... you can associate an activity with them.
- Concepts and procedures stay relevant when ...
... you can use them on your own data.

8:48



ART OF STAT Inference



Inference About Proportions

Single Population Proportion



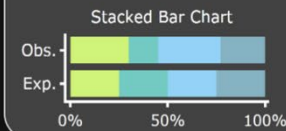
Difference of Two Population Proportions



Inference for a Population Proportion (One Sample)

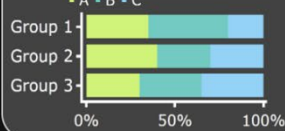
Compare Two Population Proportions (Independent Samples)

Chi-Squared Test (Goodness of Fit)



Chi-Squared Test (Goodness of Fit)

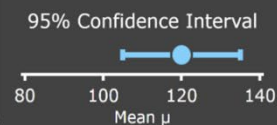
Chi-Squared Test (Independence/Homogeneity)



Chi-Squared Test (Independence or Homogeneity)

Inference About Means

Single Population Mean



Difference of Two Population Means

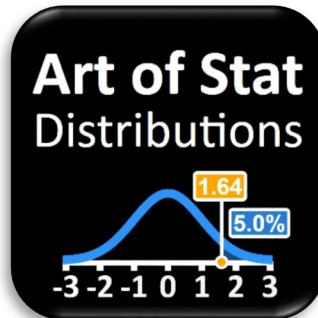
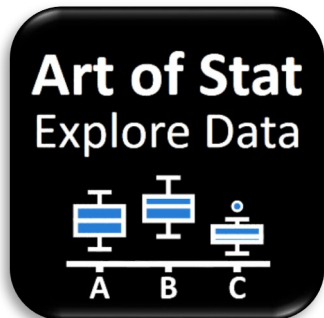


Concluding Remarks

The Art of Stat Mobile Apps provide

- ❑ A whole ecosystem of apps for (almost) the entire intro stats curriculum.
- ❑ Supports students in creating and sharing content, and telling the story.

Thank you and I'm happy to answer any questions!



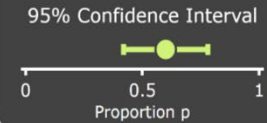
8:48



ART OF STAT Inference

Inference About Proportions

Single Population Proportion



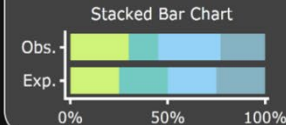
Inference for a Population Proportion (One Sample)

Difference of Two Population Proportions



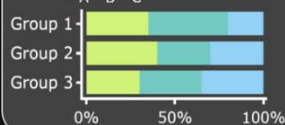
Compare Two Population Proportions (Independent Samples)

Chi-Squared Test (Goodness of Fit)



Chi-Squared Test (Goodness of Fit)

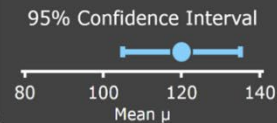
Chi-Squared Test (Independence/Homogeneity)



Chi-Squared Test (Independence or Homogeneity)

Inference About Means

Single Population Mean



Difference of Two Population Means

