

lecture 03:

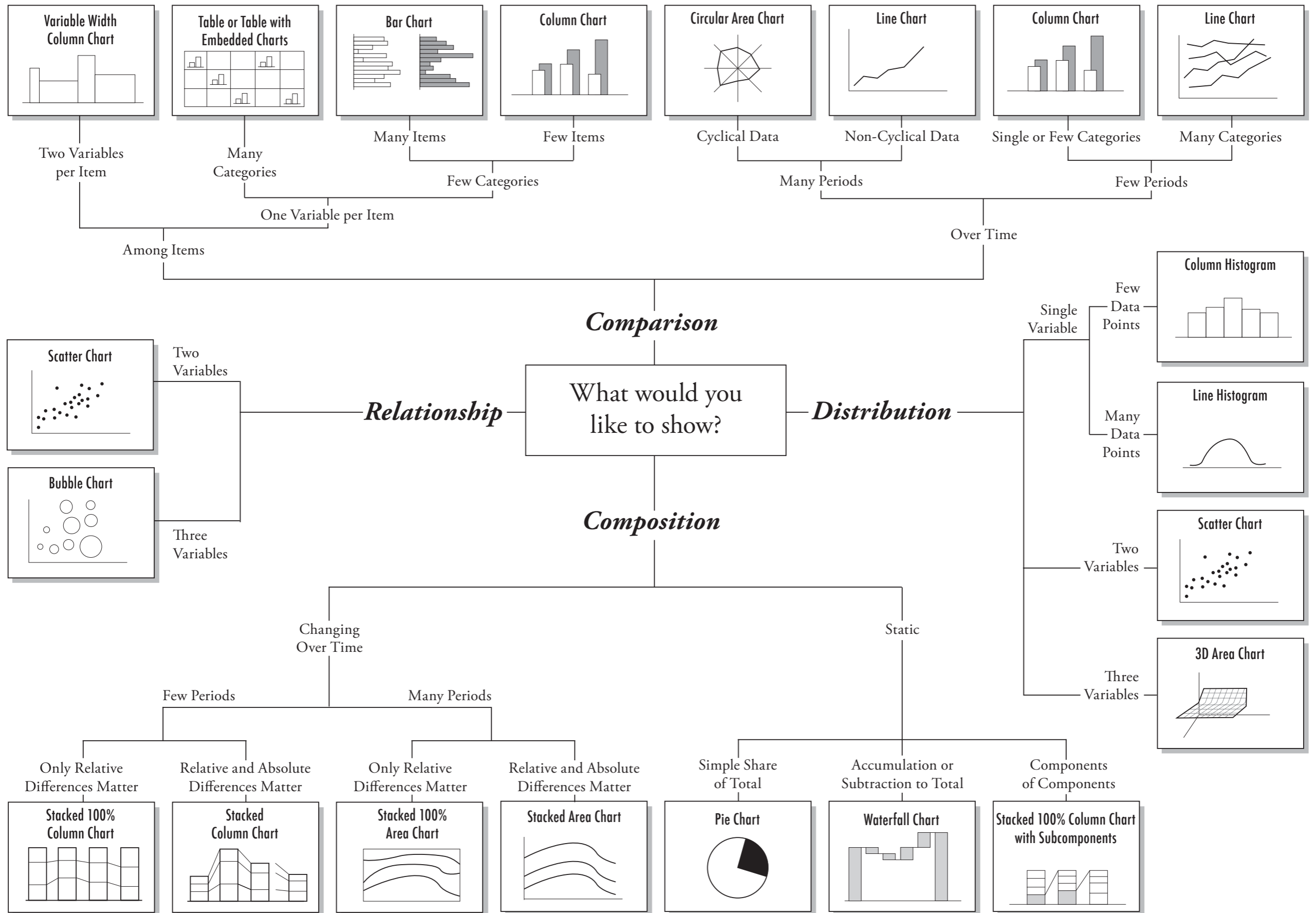
Getting Started in Plotly

September 18, 2017

But first...

Let's talk about common chart types

Chart Suggestions—A Thought-Starter



sample datasets at any time.

Portioning items into bins—the essence of a histogram

Once items are placed along a number line, drawing a histogram involves sectioning the number line into bins and counting the items that fall into each bin. Notice how the distribution shown in the histogram echoes the distribution from the dot plot.

Gathering the items into bins helps us to answer the question "what is the distribution of this data like?" Imagine trying to describe some dataset over the phone: rather than mechanically reading out the entire list of values, it would be more useful to provide a summary, such as by saying whether the variable's distribution is symmetric, where it is centered, and whether it has extreme values. A histogram is another kind of summary, in which you communicate the overall properties in terms of portions (i.e., bins) of the data.

For example, the "Geyser" data can be described as being bimodal (because its histogram has two 'peaks'), while "NBA" is more unimodal, and perhaps right-skewed (because the bin heights decrease towards the right).

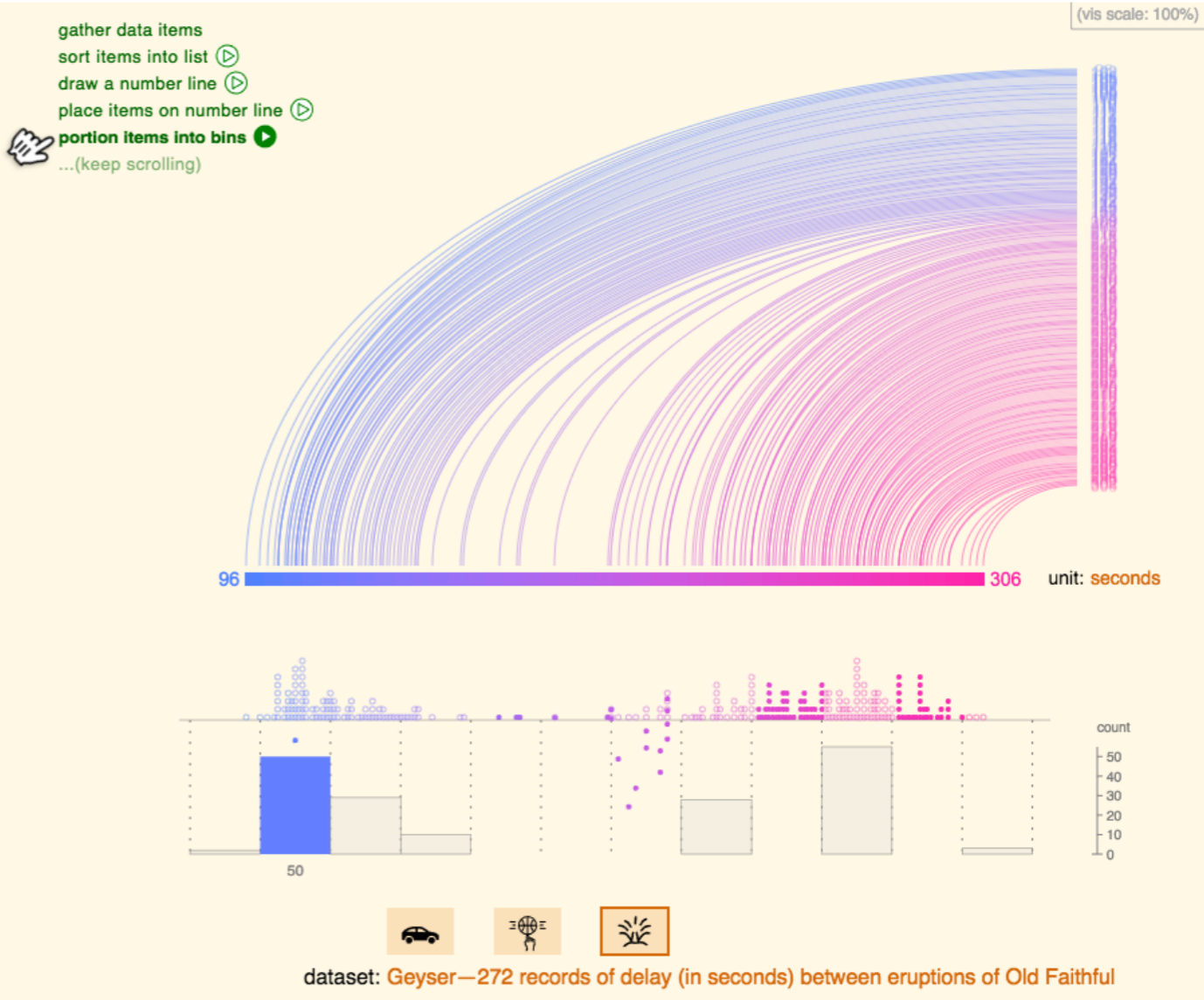
Maybe because histograms are visually similar to bar charts, it's easy to think that they are also similarly objective. But, unlike bar charts, histograms are governed by many parameters. Before describing a dataset to someone based on what you see in its histogram, you need to know whether different parameter values might have led you to different descriptions.

Bin-breaks: Why these bins?

For a start, you probably noticed that the histograms shown for our sample datasets have different numbers of bins. This is because we used Sturges' formula, a common method for estimating the number of bins for a histogram, given the size of a dataset.

Given a suggested number of bins, how did we then decide the precise values for the bin boundaries (the so-called "breaks")? Again we used a common method: look for nearby round numbers. This is why the breaks for "MPG" are all multiples of 5, and those for "NBA" are multiples of 2.

For those two datasets, the bins turn out to cover the range of the item values rather tidily. But





A company that makes data analytic software

- Browser-based tool (what we'll be using)
- API libraries for many common languages
- JavaScript library for making graphs

Step 1: Sign up for an account

<http://plot.ly>



DASH

PRICING ▾

PRODUCTS ▾

CONSULTING

PLOTCON NYC

CREATE

SIGN IN

REQUEST A DEMO

Visualize Data, Together

Plotly lets users easily create interactive charts and dashboards to share online with their audience.

TRY COMMUNITY EDITION

SEE PLANS + PRICING

Google

P&G

VTT

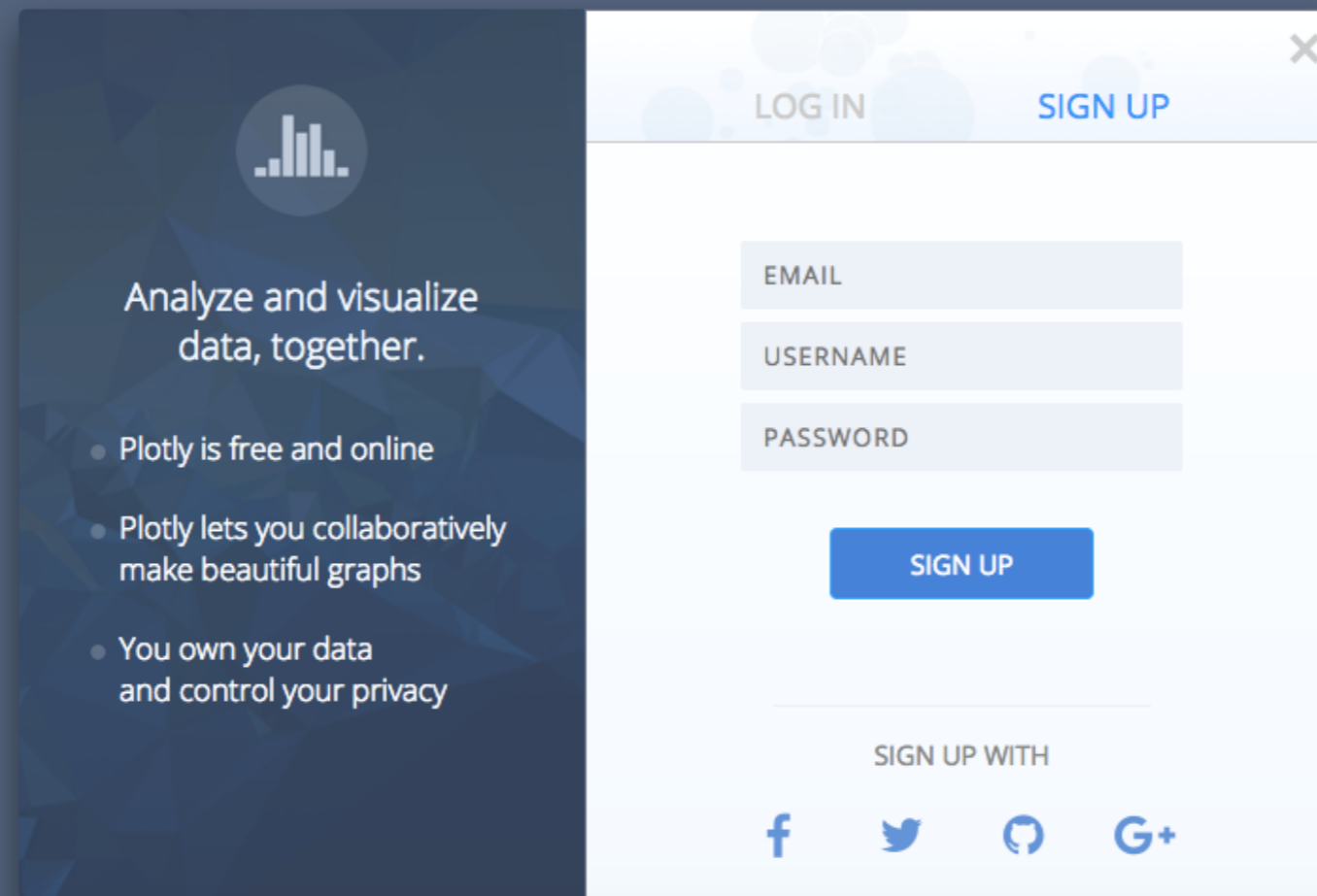
GOji

Smpl Bio
DNA



Step 1: Sign up for an account

<http://plot.ly>



The image shows a screenshot of the Plotly sign-up form. The form is divided into two main sections: a dark blue sidebar on the left and a white main content area on the right. The sidebar contains a bar chart icon, the text 'Analyze and visualize data, together.', and three bullet points: 'Plotly is free and online', 'Plotly lets you collaboratively make beautiful graphs', and 'You own your data and control your privacy'. The main content area has a 'LOG IN' link and a 'SIGN UP' link at the top. Below these are three input fields for 'EMAIL', 'USERNAME', and 'PASSWORD'. A blue 'SIGN UP' button is positioned below the input fields. At the bottom, there is a 'SIGN UP WITH' section with icons for Facebook, Twitter, GitHub, and Google+.

LOG IN SIGN UP

EMAIL

USERNAME

PASSWORD

SIGN UP

SIGN UP WITH

f t GitHub G+

Sign in and create a chart




PRICING PRODUCTS MY FILES WORKSPACE PLOTCON NYC

+ Create

AmeliaMN ▾

 Chart

 Dash App

 Dashboard

 Database Query

 Dataset

 Jupyter Notebook


 Presentation


 Folder

Plotly Community Feed

Search charts by Plotly users

Type to search

 Handpicked ▾

 Chart Type ▾



An empty chart needs data!

The screenshot displays a data visualization tool interface. On the left is a sidebar with navigation options: Graph, Create, Filter, Group, Style, Analysis, JSON, Export, Save, and Share. The main area is divided into three sections:

- Configuration Panel:** A panel titled "Collapse All" with a "+ Trace" button. It shows "Chart Type" set to "Scatter plot" with a "Try An Example" button. Below are fields for X, Y, Hover Text, Size, and Color, each with a dropdown menu.
- Data Grid:** A table labeled "Grid 1" with columns A through R and rows 1 through 4. The grid is currently empty.
- Chart Area:** A plot area with a title "Click to enter Plot title" and a y-axis labeled "Click to enter Y axis title". The y-axis has tick marks at 0.2, 0.4, 0.6, 0.8, and 1.0. The plot area is empty.

In the top right corner, there is a navigation bar with "UPGRADE" text, a blue "Import" button with an upload icon (circled in red), and a user profile "AmeliaMN" with a dropdown arrow.

Download the data file from Slack and Import it

The image shows a screenshot of a data visualization application interface. In the foreground, a white 'Import' dialog box is open, centered on the screen. The dialog has a title bar with 'Import' and a close button (X). Below the title bar, there are four tabs: 'Upload', 'By URL', 'SQL', and 'Examples'. The 'Upload' tab is selected. Inside the dialog, there is a dashed rectangular box representing a drop zone. In the center of this box is a blue 'Upload' button. Below the button, the text 'OR' is displayed, followed by 'drag and drop your files' in a larger font. At the bottom of the dialog, in smaller text, it says '(Supported file types: CSV, Excel files)'. The background is a dimmed view of the application's main interface, which includes a sidebar on the left with options like 'Graph', 'Create', 'Filter', 'Group', 'Style', 'Analysis', 'JSON', 'Export', 'Save', and 'Share'. The main area shows a grid with columns labeled L, M, N, O, P, Q, R and rows with numerical values like 0.2 and 0.4. A 'Grid 1' tab is visible at the top of the grid area. In the top right corner of the application, there are buttons for 'UPGRADE', 'Import', and a user profile 'AmeliaMN'.

Collapse All + Trace

Graph

Create

Filter

Group

Style

Analysis

JSON

Export

Save

Share

UPGRADE

Chart Type

Scatter plot

Try An Example

X

Y

Hover Text

Size

Color

UPGRADE Import AmeliaMN

Grid 1 Grid 2 +

	State	Area (Land)	Area (Water)	Region	Division	FIPS	Total Population	Urban	Rural	Male	Female	Median Age	Av
1	Alaska	1481346887193	236507354315	4	9	02	626932	411257	215675	324112	302820	32.4	
2	Arizona	294312170780	942059255	4	8	04	5130632	4523535	607097	2561057	2569575	34.2	
3	Arkansas	134855939370	2876047706	3	7	05	2673400	1404179	1269221	1304693	1368707	36	
4	California	403932815196	20036751351	4	9	06	33871648	31989663	1881985	16874892	16996756	33.3	
5	Colorado	268627156313	973947089	4	8	08	4301261	3633185	668076	2165983	2135278	34.3	
6	Connecticut	12547983374	1809174855	1	1	09	3405565	2988059	417506	1649319	1756246	37.4	
7	Delaware	5059707268	1387471795	3	5	10	783600	627758	155842	380541	403059	36	
8	District of Columbia	159033692	17968995	3	5	11	572059	572059	0	269366	302693	34.6	
9	Florida	139669812564	30633792058	3	5	12	15982378	14270020	1712358	7797715	8184663	38.7	
10													

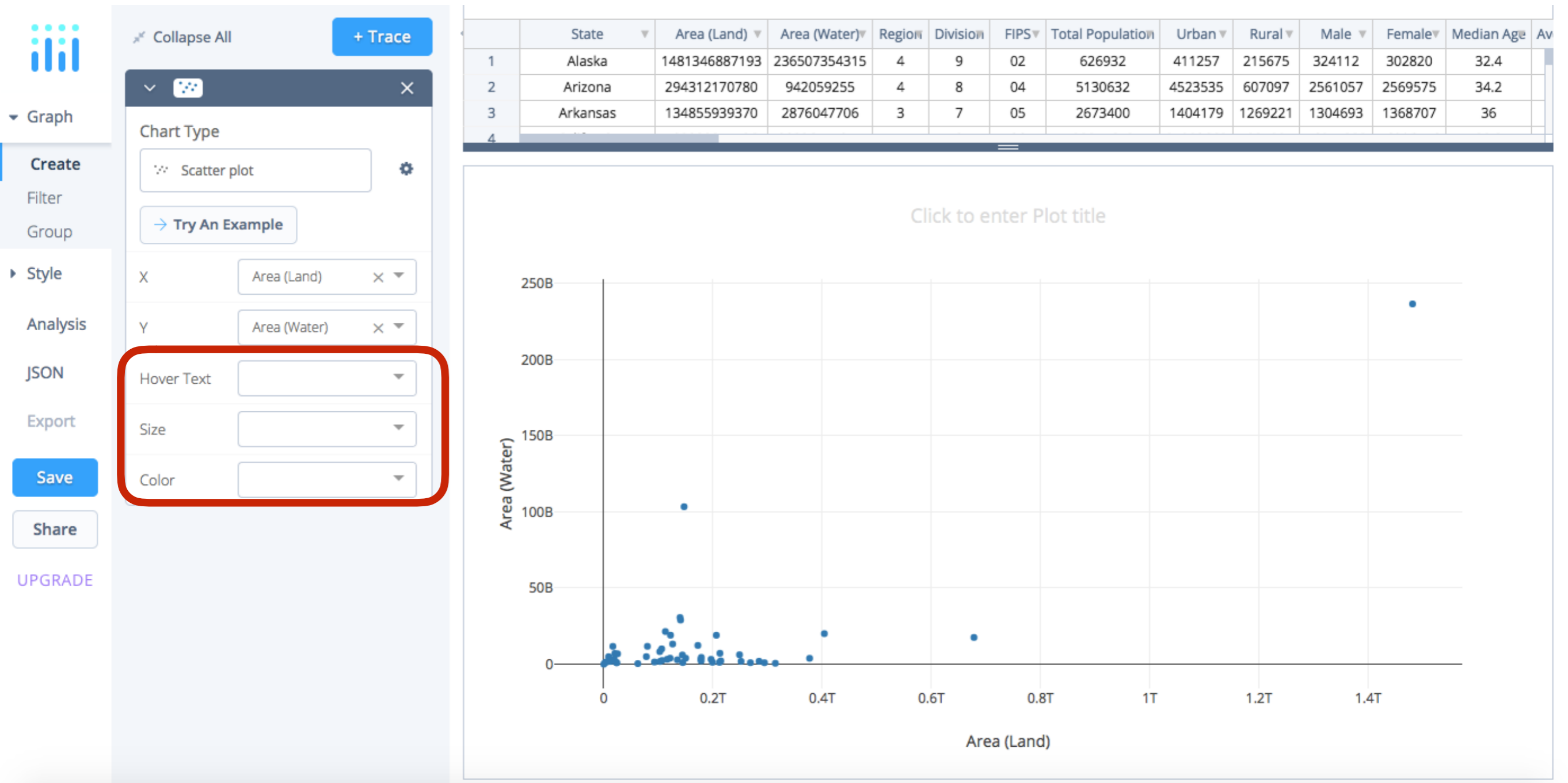
Click to enter Plot title

enter Y axis title

This is some data I grabbed from Social Explorer, which is the best way I know of to access US Census data. I found some variables that looked interesting, and cleaned them up a little.

Start simple— what are
two variables you think
might have a relationship?
Let's make a scatterplot

Try out Hover Text, Size, and Color to see what they do



It's important to save your work

Save

Your plot will be private

PLOT

filename

DATA (1)

filename

Public Private Link Private

Cancel Save UPGRADE

	State	Area (Land)	Area (Water)	Region	Division	FIPS	Total Population	Urban	Rural	Male	Female	Median Age	AV
1	Alaska	1481346887193	236507354315	4	9	02	626932	411257	215675	324112	302820	32.4	
2	Arizona	294312170780	942059255	4	8	04	5130632	4523535	607097	2561057	2569575	34.2	
3	Arkansas	134855939370	2876047706	3	7	05	2673400	1404179	1269221	1304693	1368707	36	
4													

Give your plot and data names

Choose Public privacy

You can Share your work

The image shows a data visualization interface with a scatter plot and a 'Share' dialog box. The scatter plot displays 'Area (Water)' on the y-axis (ranging from 0 to 250B) and 'Area (Land)' on the x-axis (ranging from 0 to 1.4T). A 'Share' dialog box is open, showing a shareable link, privacy settings, and data access options.

Share Dialog Box:

- Link & Privacy:** Shareable Link: <https://plot.ly/~AmeliaMN/3/>
- Privacy Settings:** Your plot will be public. Options: Public (selected), Private Link, Private.
- DATA:** demos (selected)

Data Table:

	State	Area (Land)	Area (Water)	Region	Division	FIPS	Total Population	Urban	Rural	Male	Female	Median Age	Average Household Size	PercentUrban	PercentFemale	Children	PercentChildren
1	Alaska	1481346887193	236507354315	4	9	02	626932	411257	215675	324112	302820	32.4	2.74	0.6559834240396087	0.5169811079989536	136946	0.218438363331270
2	Arizona	294312170780	942059255	4	8	04	5130632	4523535	607097	2561057	2569575	34.2	2.64	0.881672082503676	0.4991698878422775	977078	0.190440086133636
3	Arkansas	134855939370	2876047706	3	7	05	2673400	1404179	1269221	1304693	1368707	36	2.49	0.5252408917483354	0.48802760529662603	493145	0.184463604398892
4	California	403932815196	20036751351	4	9	06	33871648	31989663	1881985	16874892	16996756	33.3	2.87	0.9444377492349945	0.49820109136703356	6523949	0.192607959317479

Or, Export it

UPGRADE

Import

AmeliaMN



Graph

Create

Filter

Group

Style

Analysis

JSON

Export

Save

Share

UPGRADE

Image HTML

PNG PDF SVG EPS

Land versus water

Size As Displayed

1x (1542 x 866.5)

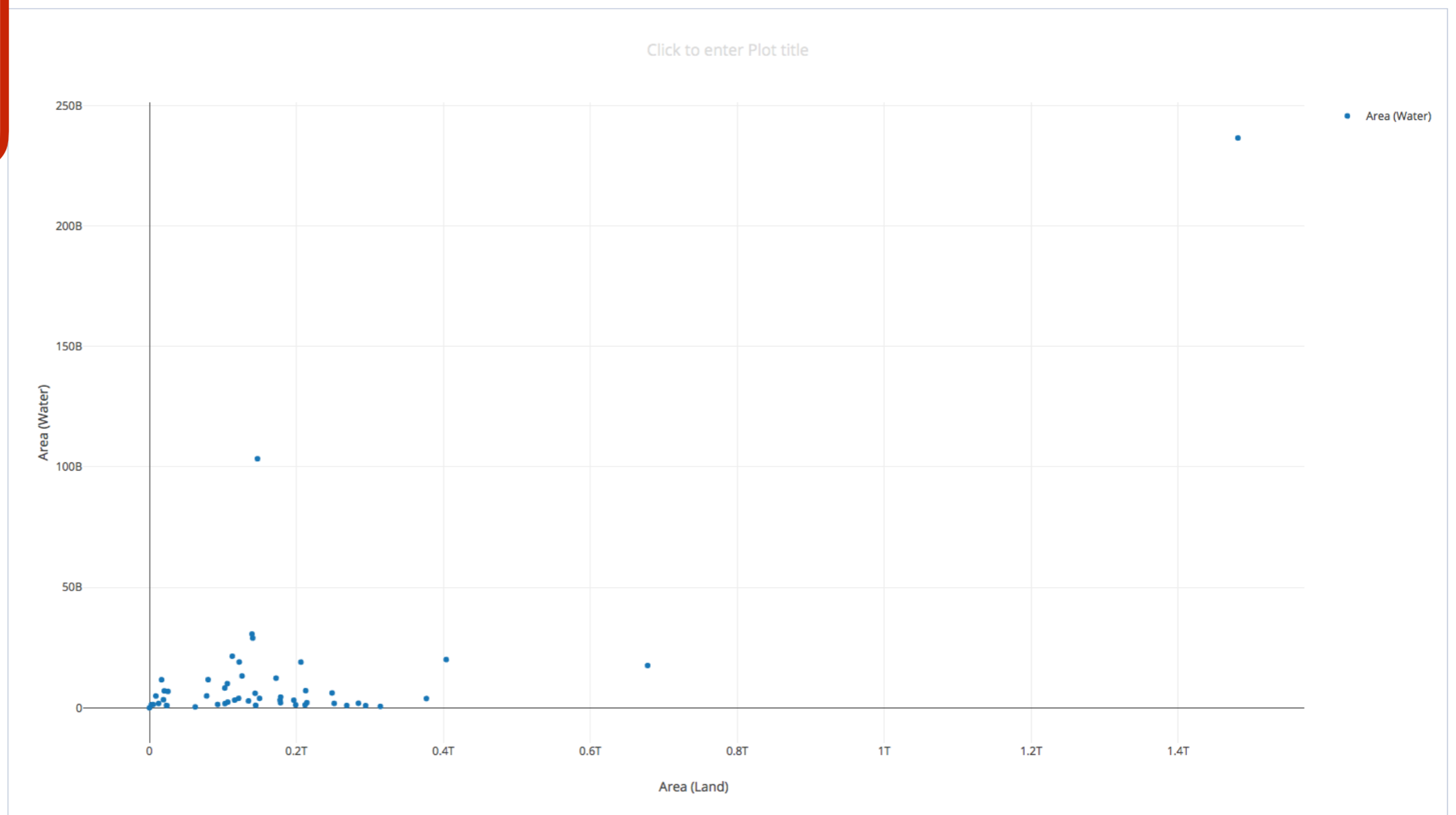
Custom Size (px)

Height 800

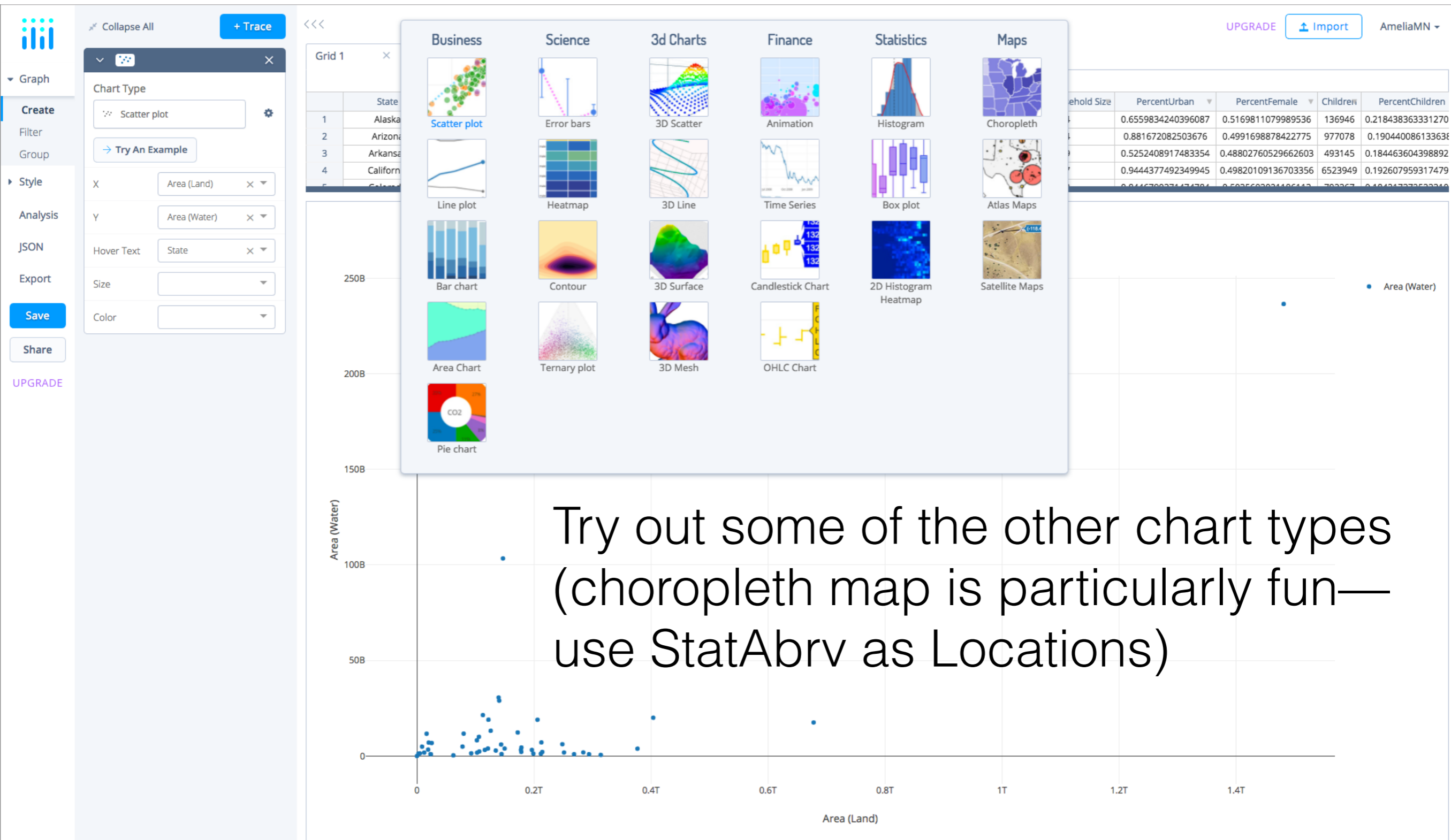
Width 1200

Download

	State	Area (Land)	Area (Water)	Region	Division	FIPS	Total Population	Urban	Rural	Male	Female	Median Age	Average Household Size	PercentUrban	PercentFemale	Children	PercentChildren
1	Alaska	1481346887193	236507354315	4	9	02	626932	411257	215675	324112	302820	32.4	2.74	0.6559834240396087	0.5169811079989536	136946	0.218438363331270
2	Arizona	294312170780	942059255	4	8	04	5130632	4523535	607097	2561057	2569575	34.2	2.64	0.881672082503676	0.4991698878422775	977078	0.190440086133638
3	Arkansas	134855939370	2876047706	3	7	05	2673400	1404179	1269221	1304693	1368707	36	2.49	0.5252408917483354	0.48802760529662603	493145	0.184463604398892
4	California	403932815196	20036751351	4	9	06	33871648	31989663	1881985	16874892	16996756	33.3	2.87	0.9444377492349945	0.49820109136703356	6523949	0.192607959317479



Now it's time to play!



Lab assignment:

- Make a plot that you think is interesting (remember to save!)
- Use the Share button to get the link to your visualization
- Post the link on Slack in the channel #lab1, along with a written description of what you think the visualization shows
- Post any questions you have about Plotly in the #questions channel