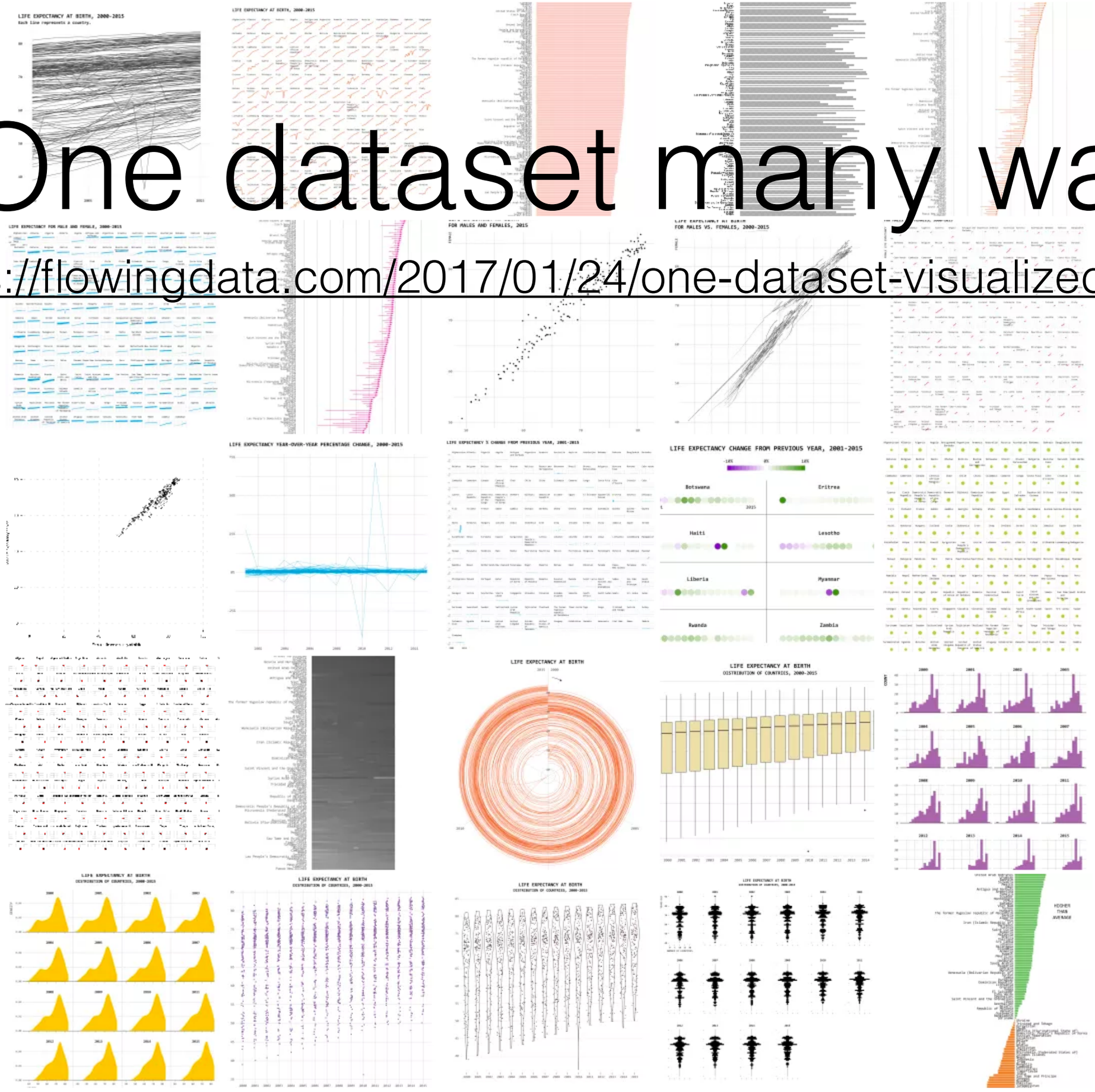
lecture 05:

Aggregating data and 1-D charts

September 27, 2017

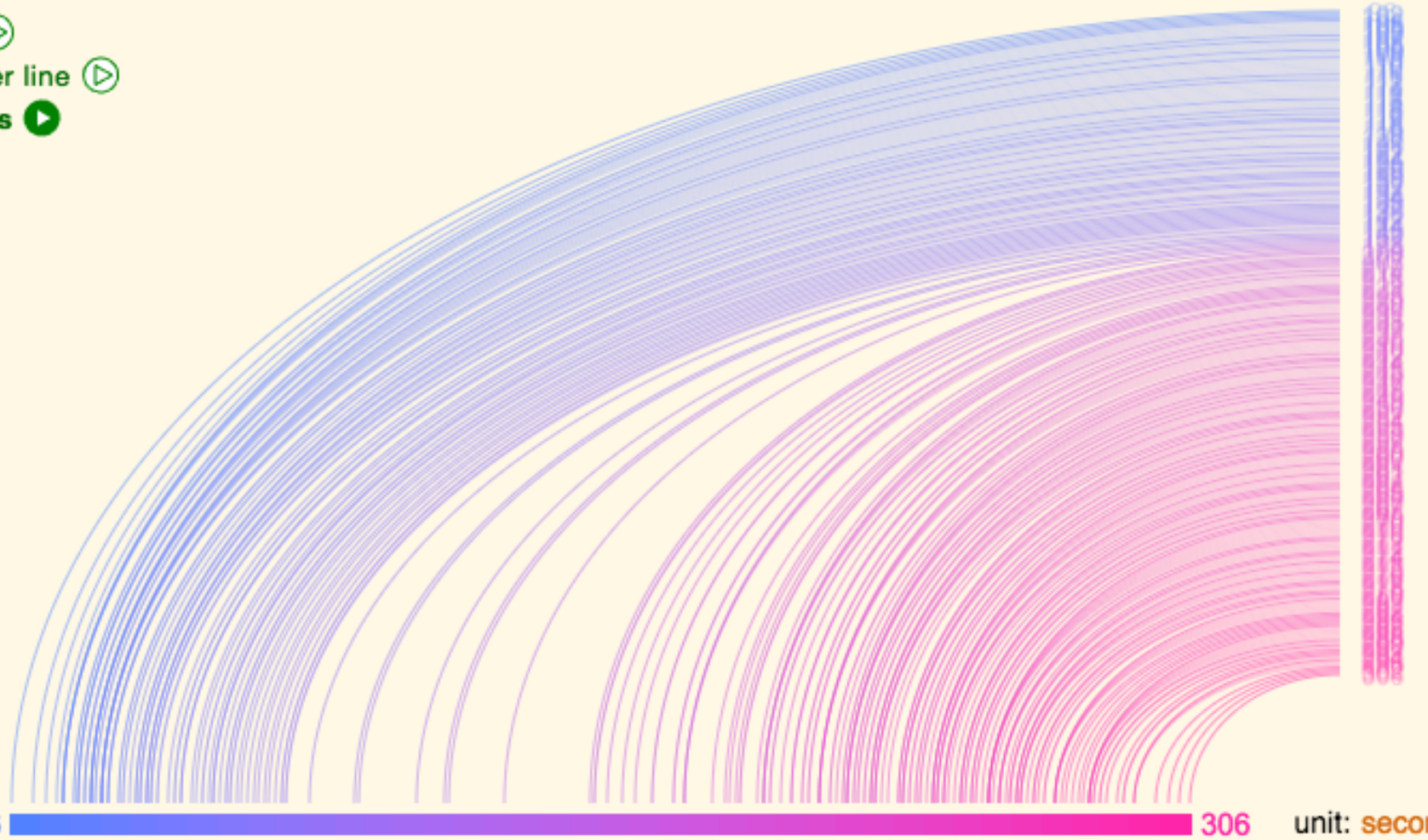
One dataset many ways

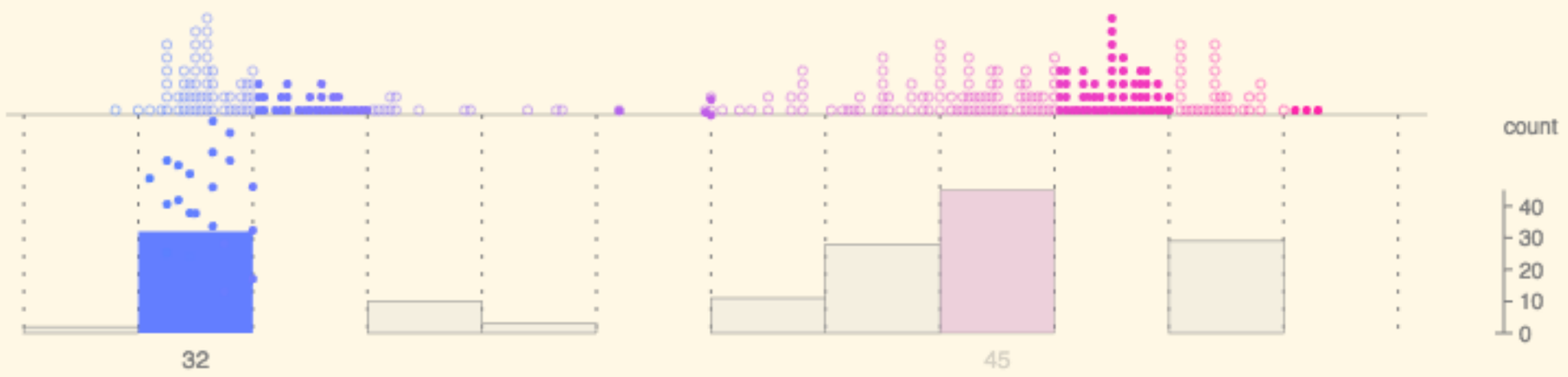
https://flowingdata.com/2017/01/24/one-dataset-visualized-25-ways/



gather data items
sort items into list
draw a number line
place items on number line
portion items into bins
...(keep scrolling)

(vis scale: 100%)





dataset: Geyser—272 records of delay (in seconds) between eruptions of Old Faithful

Histograms in Plotly

Collapse All

+ Trace

Graph

Create

Filter

Group

Style

Traces

Layout

Notes

Axes

Legend

Shapes

Images

Mobile

Analysis

JSON

Export

Save

Share

UPGRADE

Chart Type

Histogram

Values

Average Househ...

<<<

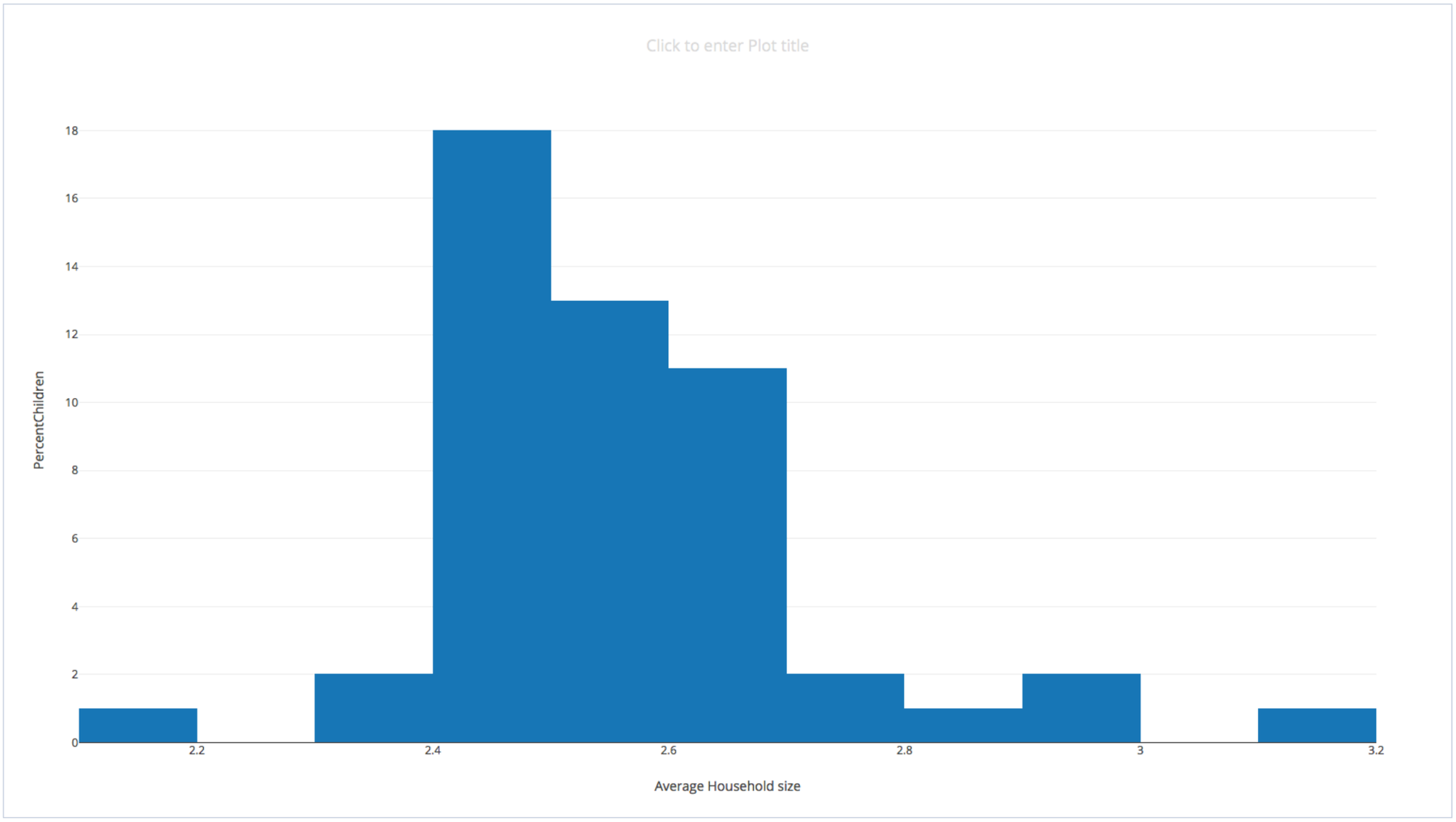
Grid 1

UPGRADE

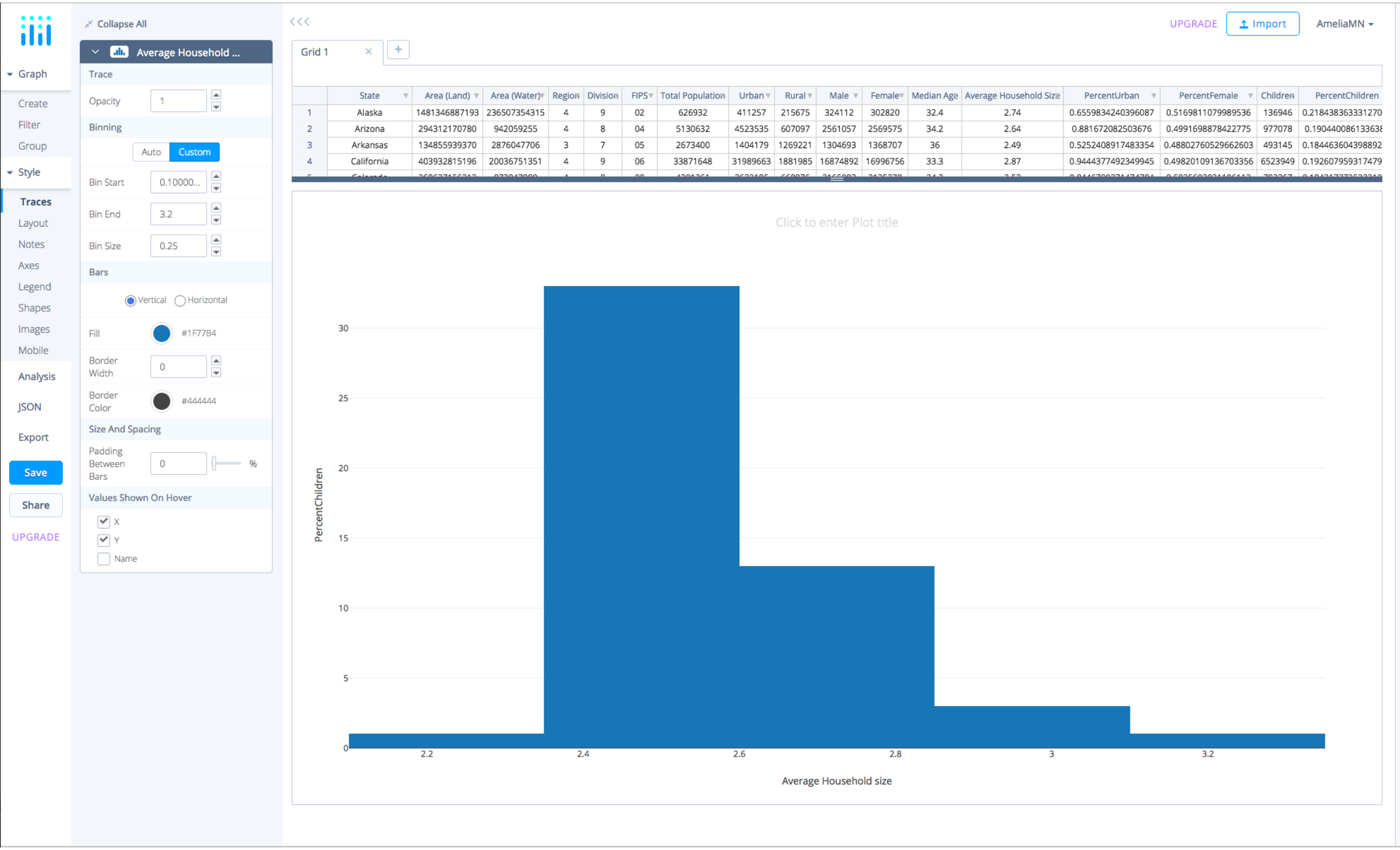
Import

AmeliaMN

	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



https://plot.ly



	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

Pivot tables

- A way to aggregate data in tools like Excel and Google Sheets
- Often thought of as a “magic” skill in business
- Not actually that complicated!
- Like

```
demos %>%
  group_by(region) %>%
  summarize(total=n())
```

in R

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

Data menu:
- Sort sheet by **column A**, A → Z
- Sort sheet by **column A**, Z → A
- Sort range by **column A**, A → Z
- Sort range by **column A**, Z → A
- Sort range...
- Randomize range
- Named ranges...
- Protected sheets and ranges...
- Split text to columns...
- Filter
- Filter views...
- Pivot table...
- Data validation...

fx State

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	State	Area (Land)	Area			FIPS	Total Population	Urban	Rural	Male	Female	Median Age	Average Househ	PercentUrban	PercentFemale	Children	PercentChildren	RetirementAge	PercentRetireme
2	Connecticut	12547983374	1		1	9	3405565	2988059	417506	1649319	1756246	37.4	2.53	0.8774047772	0.4843011365	597544	0.1754610468	470183	0.1380631408
3	Maine	79931052601	11		1	23	1274923	512878	762045	620309	654614	38.6	2.39	0.4022815496	0.4865462463	218216	0.1711601407	183402	0.1438533935
4	Massachusetts	20305551568	7		1	25	6349097	5801367	547730	3058816	3290281	36.5	2.51	0.9137310392	0.4817718173	1069203	0.1684023728	860162	0.1354778483
5	New Hampshire	23227265747			1	33	1235786	732335	503451	607687	628099	37.1	2.53	0.5926066487	0.4917412885	221025	0.1788537821	147970	0.1197375597
6	New Jersey	19210817801	3		2	34	8414350	7939087	475263	4082813	4331537	36.7	2.68	0.9435175623	0.4852202487	1483029	0.1762499777	1113136	0.1322901947
7	New York	122283145776	19		2	36	18976457	16602582	2373875	9146748	9829709	35.9	2.61	0.8749042037	0.4820050445	3338250	0.1759153461	2448352	0.1290205016
8	Pennsylvania	116074476061	3		2	42	12281054	9464101	2816953	5929663	6351391	38	2.48	0.7706261205	0.4828301382	2094276	0.1705290116	1919165	0.156270382
9	Rhode Island	2706348685	1		1	44	1048319	953146	95173	503635	544684	36.7	2.47	0.9092137031	0.4804215129	175917	0.1678086537	152402	0.1453775044
10	Vermont	23956246818			1	50	608827	232448	376379	298337	310490	37.7	2.44	0.3817964709	0.4900193323	106422	0.1747984239	77510	0.1273103854
11	Illinois	143960815935	6		3	17	12419293	10909520	1509773	6080336	6338957	34.7	2.63	0.878433257	0.4895879339	2315593	0.1864512738	1500025	0.1207818352
12	Indiana	92894848176	1		3	18	6080485	4304011	1776474	2982474	3098011	35.2	2.53	0.7078400818	0.4904993598	1131123	0.1860251279	752831	0.1238110118
13	Iowa	144700963406	1		4	19	2926324	1787432	1138892	1435515	1490809	36.6	2.46	0.6108113797	0.4905523107	531035	0.1814682858	436213	0.1490651753
14	Kansas	211899555547	1		4	20	2688418	1920669	767749	1328474	1359944	35.2	2.51	0.7144235011	0.4941471155	517419	0.1924622585	356229	0.1325050643
15	Michigan	147121215368	103		3	26	9938444	7419457	2518987	4873095	5065349	35.5	2.56	0.7465411084	0.4903277616	1850586	0.1862048023	1219018	0.1226568264
16	Minnesota	206189140788	1898 1482285	2	4	27	4919479	3490059	1429420	2435631	2483848	35.4	2.52	0.7094367107	0.4950993794	931000	0.1892476825	594266	0.1207985642
17	Missouri	178413721374	2119611899	2	4	29	5595211	3883442	1711769	2720177	2875034	36.1	2.48	0.6940653355	0.4861616479	1028794	0.1838704564	755379	0.1350045601
18	Nebraska	199098632058	1246588932	2	4	31	1711263	1193725	517538	843351	867912	35.3	2.49	0.6975695729	0.4928237214	326797	0.1909683082	232195	0.1356863323
19	North Dakota	178646833011	4464770063	2	4	38	642200	358958	283242	320524	321676	36.2	2.41	0.5589504827	0.4991030832	117867	0.1835362815	94478	0.1471161632
20	Ohio	106055800423	10040156786	2	3	39	11353140	8782329	2570811	5512262	5840878	36.2	2.49	0.7735594734	0.4855275281	2071993	0.1825039593	1507757	0.1328052856
21	South Dakota	196540314429	3190480713	2	4	46	754844	391427	363417	374558	380286	35.6	2.5	0.5185535024	0.4962058386	148163	0.1962829406	108131	0.1432494661
22	Wisconsin	140662519921	28976054385	2	3	55	5363675	3663643	1700032	2649041	2714634	36	2.5	0.6830471645	0.4938854424	989272	0.1844392138	702553	0.1309835141
23	Arkansas	134855939370	2876047706	3	7	5	2673400	1404179	1269221	1304693	1368707	36	2.49	0.5252408917	0.4880276053	493145	0.1844636044	374019	0.1399038677
24	Delaware	5059707268	1387471795	3	5	10	783600	627758	155842	380541	403059	36	2.54	0.8011204696	0.4856316998	138774	0.1770980092	101726	0.1298187851
25	District of Columl	159033692	17968995	3	5	11	572059	572059	0	269366	302693	34.6	2.16	1	0.4708710116	79607	0.1391587231	69898	0.1221866975
26	Florida	139669812564	30633792058	3	5	12	15982378	14270020	1712358	7797715	8184663	38.7	2.46	0.8928596233	0.4878945424	2614622	0.1635940534	2807597	0.1756682892
27	Georgia	149976204860	3933225333	3	5	13	8186453	5864163	2322290	4027113	4159340	33.4	2.65	0.716325251	0.491924036	1553650	0.1897830477	785275	0.0959237169
28	Kentucky	102895503258	1763380947	3	6	21	4041769	2253800	1787969	1975368	2066401	35.9	2.47	0.5576271182	0.4887384707	715560	0.1770412906	504793	0.1248940748
29	Louisiana	112824665700	21439542252	3	7	22	4468976	3245665	1223311	2162903	2306073	34	2.62	0.7262659276	0.4839817891	883019	0.1975886646	516929	0.1156705697
30	Maryland	25314076865	6819073655	3	5	24	5296486	4558668	737818	2557794	2738692	36	2.61	0.8606966959	0.4829228285	964854	0.1821687058	599307	0.1131518142
31	Mississippi	121488450732	3945172740	3	6	28	2844658	1387351	1457307	1373554	1471104	33.8	2.63	0.48770397	0.4828538264	558267	0.196251008	343523	0.1207607382
32	North Carolina	126160600432	13228705076	3	5	37	8049313	4849482	3199831	3942695	4106618	35.3	2.49	0.6024715401	0.4898175782	1401494	0.1741134927	969048	0.1203889077
33	Oklahoma	177846860194	3188623148	3	7	40	3450654	2254563	1196091	1695895	1754759	35.5	2.49	0.653372665	0.4914706024	647835	0.1877426714	455950	0.1321343722
34	South Carolina	77983161188	4948771376	3	5	45	4012012	2427124	1584888	1948929	2063083	35.4	2.53	0.6049642922	0.4857734723	724398	0.1805572865	485333	0.1209699772
35	Tennessee	106751842839	2398721489	3	6	47	5689283	3620018	2069265	2770275	2919008	35.9	2.48	0.6362872088	0.4869286692	1002708	0.1762450558	703311	0.1236203226
36	Texas	678051389889	17569713726	3	7	48	20851820	17204281	3647539	10352910	10498910	32.3	2.74	0.8250733509	0.4964991066	4232575	0.2029834806	2072532	0.09939333833
37	Virginia	102548167005	8236501395	3	5	51	7078515	5169955	1908560	3471895	3606620	35.7	2.54	0.7303728254	0.4904835266	1243178	0.17562695	792333	0.1119349186
38	West Virginia	62361027006	393757137	3	5	54	1808344	832780	975564	879170	929174	38.9	2.4	0.4605207859	0.4861740908	291243	0.1610550869	276895	0.1531207558
39	Alaska	1481346887193	236507354315	4	9	2	626932	411257	215675	324112	302820	32.4	2.74	0.655983424	0.516981108	136946	0.2184383633	35699	0.05694237972
40	Arizona	294312170780	942059255	4	8	4	5130632	4523535	607097	2561057	2569575	34.2	2.64	0.8816720825	0.4991698878	977078	0.1904400861	667839	0.1301670048
41	California	403932815196	20036751351	4	9	6	33871648	31989663	1881985	16874892	16996756	33.3	2.87	0.9444377492	0.4982010914	6523949	0.1926079593	3595658	0.1061553899
42	Colorado	268627156313	973947089	4	8	8	4301261	3633185	668076	2165983	2135278	34.3	2.53	0.8446790371	0.5035693021	792367	0.1842173725	416073	0.09673279534
43	Hawaii	16634521307	11676592338	4	9	15	1211537	1108225	103312	608671	602866	36.2	2.92	0.9147265003	0.5023957172	210787	0.1739831305	160601	0.1325597155
44	Idaho	214314284702	2131225477	4	8	16	1293953	859497	434456	648660	645293	33.2	2.69	0.6642412823	0.5013010519	268274	0.2073290143	145916	0.1127676198
45	Montana	376979057369	3858989004	4	8	30	902195	487878	414317	449480	452715	37.5	2.45	0.5407677941	0.4982071503	168099	0.1863222474	120949	0.1340608183
46	Nevada	284448008848	1902899301	4	8	32	1998257	1828646	169611	1018051	980206	35	2.62	0.9151205275	0.5094695027	362477	0.1813965871	218929	0.1095599815
47	New Mexico	314309355994	605947499	4	8	35	1819046	1363501	455545	894317	924729	34.6	2.63	0.7495692797	0.4916406732	367403	0.201975651	212225	0.1166682976

demos_cleaned.csv | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Sum: 9705584287944 Explore

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

100% $ % .0 .00 123 Arial 10

Report Editor

demos_cleaned.csv!1:104 -
Edit range...

Rows - Add field

Columns - Add field

Values - Add field

Filter - Add field

demos_cleaned.csv Pivot Table 5 Pivot Table 4 Pivot Table 2 Pivot Table 1

Explore

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

100% $ % .0 .00 123 Arial 10

Report Editor

demos_cleaned.csv!1:104 -
Edit range...

Rows - Add field

State
Area (Land)
Area (Water)
Region
Division
FIPS
Total Population
Urban
Rural
Male
Female
Median Age
Average Household Size
PercentUrban
PercentFemale
Children
PercentChildren
RetirementAge
PercentRetirementAge
PercentWhite
PercentBlack
PercentNativeAmerican
PercentAsian
PercentNativeHawaii
PercentOther
PercentTwoOrMoreRaces
StateAbrv
Column AB
Column AC

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

100% $ % .0 .00 123 Arial 10

fx

	A	B
1		
2	1	
3	2	
4	3	
5	4	
6	9	
7	Grand Total	

Report Editor

demos_cleaned.csv!1:104 -
Edit range...

Rows - Add field

Group by: Region

Order: Ascending

Sort by: Region

Show totals

Columns - Add field

Values - Add field

State
Area (Land)
Area (Water)
Region
Division
FIPS
Total Population
Urban
Rural
Male
Female
Median Age
Average Household Size
PercentUrban
PercentFemale
Children
PercentChildren
RetirementAge
PercentRetirementAge
PercentWhite
PercentBlack
PercentNativeAmerican
PercentAsian
PercentNativeHawaii
PercentOther
PercentTwoOrMoreRaces
StateAbrv
Column AB

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

100% $ % .0 .00 123 Arial 10

	A	B
1		0
2	1	0
3	2	0
4	3	0
5	4	0
6	9	0
7	Grand Total	0

Report Editor

demos_cleaned.csv!1:104 -
Edit range...

Rows - Add field

Group by: Region
Order: Ascending
Sort by: Region
Show totals

Columns - Add field

Values - Add field

Display: State
Summarize by: SUM

SUM
COUNTA
COUNT
COUNTUNIQUE
AVERAGE
MAX
MIN
MEDIAN
PRODUCT
STDEV
STDEVP
VAR
VARP

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

100% $ % .0 .00 123 Arial 10

fx

	A	B
1		0
2	**1**	9
3	**2**	12
4	**3**	16
5	**4**	13
6	**9**	1
7	**Grand Total**	51

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

Now, let's make a bar chart in Plotly!

- Copy and paste your pivot table data into Plotly
- Create a bar chart (Y will be the counts, and X will be the groups)
- Change the color of your bar chart and adjust the width between the bars
- Annotate your bar chart with what each of the groups means. You may need to look back at Google sheets

More complicated pivot tables

- Allow you to summarize another quantitative variable
- Like

```
demos %>%
  group_by(region) %>%
  summarize(average=mean(PercUrban))
```

in R

- Allow you to summarize two categorical variables
- Like

```
demos %>%
  group_by(region, division) %>%
  summarize(total=n())
```

in R

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

	A	B
1	1	
2	2	
3	3	
4	4	
5	9	
6	**Grand Total**	

Report Editor

demos_cleaned.csv!1:52 -
Edit range...

Rows - Add field

Group by: Region
Order: Ascending
Sort by: Region
Show totals

Columns - Add field

Values - Add field

Region
Division
FIPS
Total Population
Urban
Rural
Male
Female
Median Age
Average Household Size
PercentUrban
PercentFemale
Children
PercentChildren
RetirementAge
PercentRetirementAge
PercentWhite
PercentBlack
PercentNativeAmerican
PercentAsian
PercentNativeHawaii
PercentOther
PercentTwoOrMoreRaces
StateAbrv
Column AB
Column AC

Calculated Field

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

	A	B
1	1	6.666082075
2	2	8.29323157
3	3	11.08090262
4	4	10.25137484
5	9	0.9440507167
6	**Grand Total**	37.23564182

Report Editor

demos_cleaned.csv!1:52 -
Edit range...

Rows - Add field

Group by: Region
Order: Ascending
Sort by: Region
Show totals

Columns - Add field

Values - Add field

Display: PercentUrban
Summarize by: SUM

SUM
COUNTA
COUNT
COUNTUNIQUE
AVERAGE
MAX
MIN
MEDIAN
PRODUCT
STDEV
STDEVP
VAR
VARP

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Sum: 56.23564182

Explore

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

100% $ % .0 .00 123 Arial 10

1	**1**	0.7406757861
2	**2**	0.6911026309
3	**3**	0.6925564135
4	**4**	0.7885672955
5	**9**	0.9440507167
6	**Grand Total**	0.7301106239

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

SDS%2FMTH 220 Fall ....csv

Show All

Making a bar chart in Plotly is just the same

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

100% $ % .0 .00 123 Arial 10

	A	B
1		
2	1	
3	2	
4	3	
5	4	
6	9	

Report Editor

demos_cleaned.csv!1:104 -
Edit range...

Rows - Add field

- State
- Area (Land)
- Area (Water)
- Division
- FIPS
- Total Population
- Urban
- Rural
- Male
- Female
- Median Age
- Average Household Size
- PercentUrban
- PercentFemale
- Children
- PercentChildren
- RetirementAge
- PercentRetirementAge
- PercentWhite
- PercentBlack
- PercentNativeAmerican
- PercentAsian
- PercentNativeHawaii
- PercentOther
- PercentTwoOrMoreRaces
- StateAbrv
- Column AB
- Column AC

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

Demos_cleaned

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

amcnamara@smith.edu

Comments Share

	A	B	C
1			
2	1	1	
3		2	
4	2	3	
5		4	
6	3	5	
7		6	
8		7	
9	4	8	
10		9	
11	9	0	

Report Editor

demos_cleaned.csv!1:104 -
Edit range...

Rows - Add field

Group by: Region
Order: Ascending
Sort by: Region
Show totals

Group by: Division
Order: Ascending
Sort by: Division
Show totals

Columns - Add field

Values - Add field

- State
- Area (Land)
- Area (Water)
- Region
- Division
- FIPS
- Total Population
- Urban
- Rural
- Male
- Female
- Median Age
- Average Household Size
- PercentUrban
- PercentFemale
- Children
- PercentChildren
- RetirementAge
- PercentRetirementAge
- PercentWhite
- PercentBlack
- PercentNativeAmerican
- PercentAsian

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

Demos_cleaned

amcnamara@smith.edu

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

100% $ % .0 .00 123 Arial 10

		#DIV/0!	
1	1	1870038.833	
	2	11335256.67	
2	3	7015792	
	4	1860816	
3	5	4352445.444	
	6	2420389.667	
	7	6027172	
4	8	1873491.25	
	9	8206879	
9	0	3595521	

demos_cleaned.csv | Pivot Table 5 | Pivot Table 4 | Pivot Table 2 | Pivot Table 1

Explore

Making a bar chart in Plotly is a little more complicated

Collapse All + Trace

Chart Type

Bar chart

Y C

X B

Hover Text C

Graph

Create

Filter

Group

Style

Traces

Layout

Notes

Axes

Legend

Shapes

Images

Mobile

Analysis

JSON

Export

Save

Share

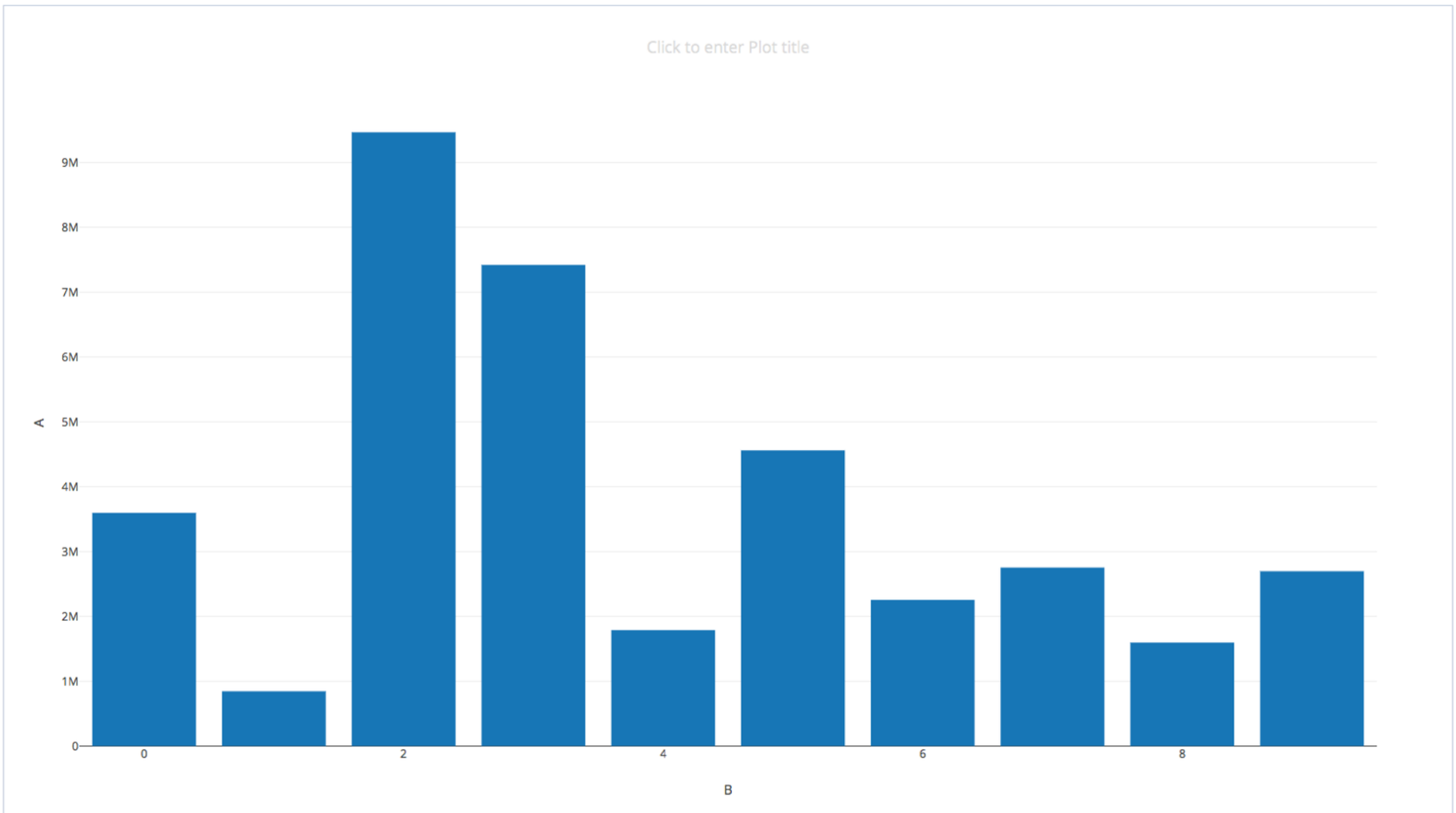
UPGRADE

<<<

UPGRADE Import AmeliaMN

Grid 1

	A	B	C
1	1	1	842740.5
2	1	2	9464101
3	2	3	7419457
4	2	4	1787432



Collapse All

+ Group

Grouping on A

Group By: A

Across Traces: C

Graph

Create

Filter

Group

Style

Traces

Layout

Notes

Axes

Legend

Shapes

Images

Mobile

Analysis

JSON

Export

Save

Share

UPGRADE

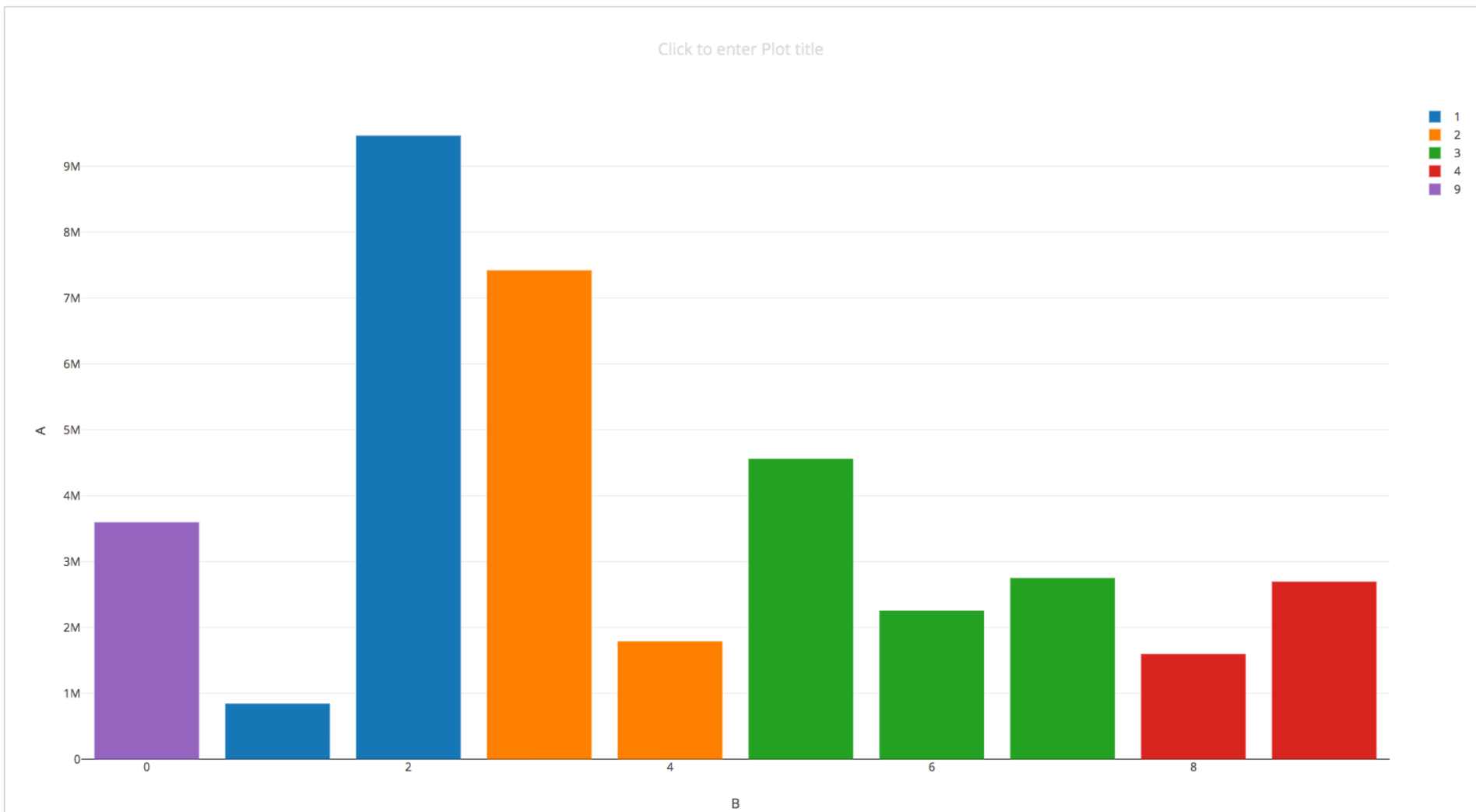
UPGRADE

Import

AmeliaMN

Grid 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	A
1	1	1	842740.5																											
2	1	2	9464101																											
3	2	3	7419457																											
4	2	4	1787432																											



Your lab assignment:

- Create one bar chart (either count or value) or histogram from this dataset
- Post screenshot in #lab2
- Write something about your design decisions