Today's Agenda

Exploring bivariate and multivariate visualizations

Justin Leinaweaver (Spring 2022)

Dataset 1: The Motivating Problem

What drives economic investment in US states?

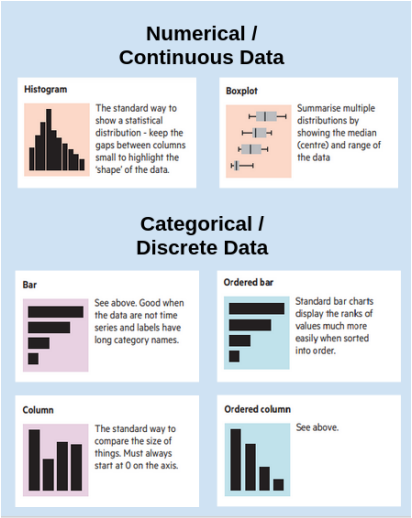
Why do some states attract greater investment by companies and individuals than others?

Last Week: Univariate Analyses

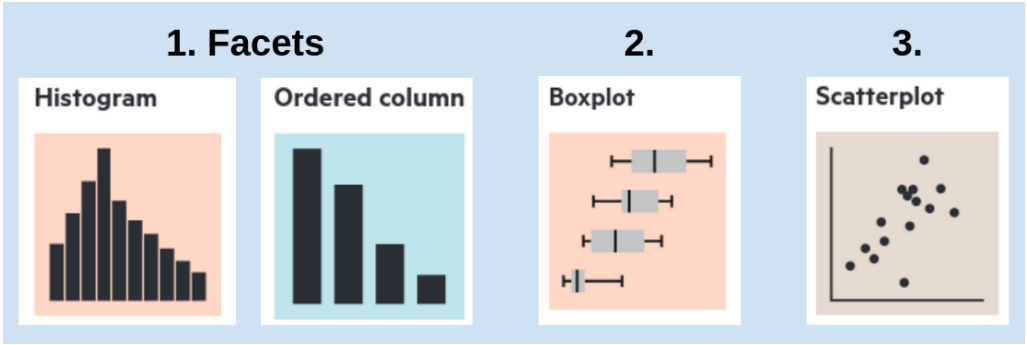
Measures of Central Tendency
- Mean
- Median

Deviations from Central Tendency
- Standard deviation

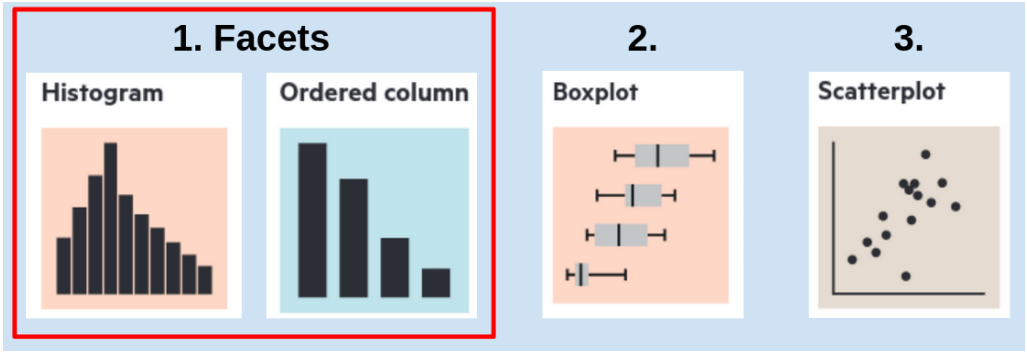
Measures of Variability
- Range
- IQR



Bivariate and Multivariate Visualizations



Bivariate and Multivariate Visualizations



Subset the data and then make your bar plot or histogram

manufacturing_thousands and income_tax

1. Copy the data to a new sheet

	A	B	C
1	State	manufacturing_thousands	income_tax
2	Alabama	257.8	1
3	Alaska	11.6	0
4	Arizona	176.4	1
5	Arkansas	155.1	1
6	California	1261.7	1
7	Colorado	146.5	1
8	Connecticut	153.6	1
9	Delaware	25.6	1
10	Florida	376.7	0
11	Georgia	385.3	1
12	Hawaii	11.9	1
13	Idaho	68.2	1

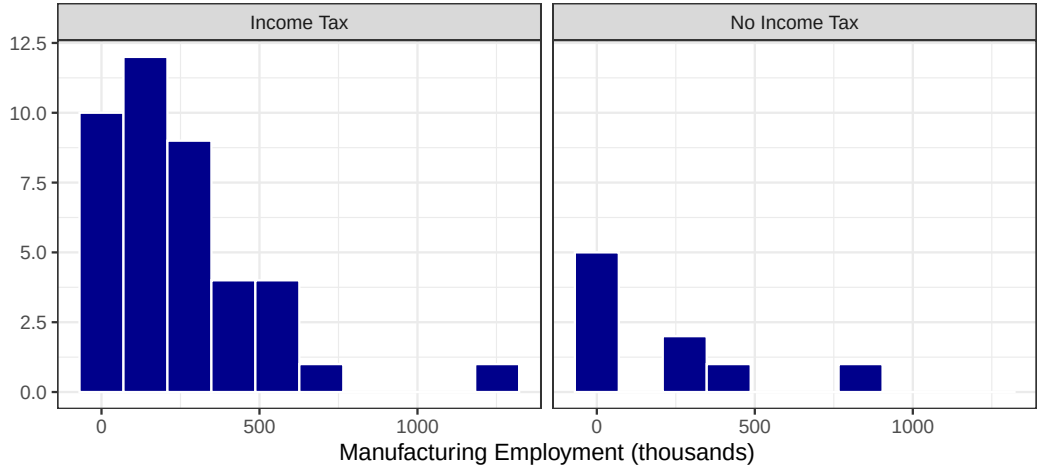
2. Sort the data by the categorical variable

	A	B	C
1	State	manufacturing_thousands	income_tax
2	Alaska	11.6	0
3	Florida	376.7	0
4	Nevada	56.1	0
5	New Hampshi	67.3	0
6	South Dakota	43.2	0
7	Tennessee	335.1	0
8	Texas	868.7	0
9	Washington	272	0
10	Wyoming	9.5	0
11	Alabama	257.8	1
12	Arizona	176.4	1
13	Arkansas	155.1	1
14	California	1261.7	1
15	Colorado	146.5	1

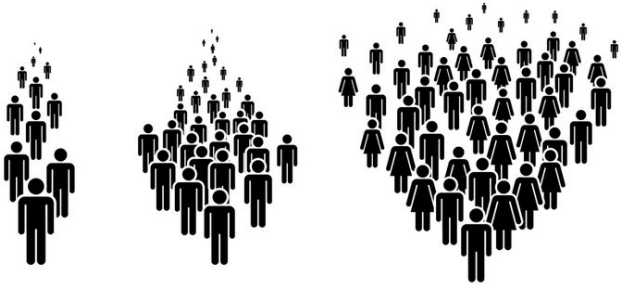
3. Split the data by subset

	A	B	C	D	E	F	G
1	State	manufacturing_thousands	income_tax		State	manufacturing_thousands	income_tax
2	Alaska	11.6	0		Alabama	257.8	1
3	Florida	376.7	0		Arizona	176.4	1
4	Nevada	56.1	0		Arkansas	155.1	1
5	New Hampshi	67.3	0		California	1261.7	1
6	South Dakota	43.2	0		Colorado	146.5	1
7	Tennessee	335.1	0		Connecticut	153.6	1
8	Texas	868.7	0		Delaware	25.6	1
9	Washington	272	0		Georgia	385.3	1
10	Wyoming	9.5	0		Hawaii	11.9	1
11					Idaho	68.2	1
12					Illinois	554.2	1
13					Indiana	505.1	1
14					Iowa	216.6	1
15					Kansas	159	1

4. Make a histogram for each subset



Does population size predict the size of your manufacturing sector?



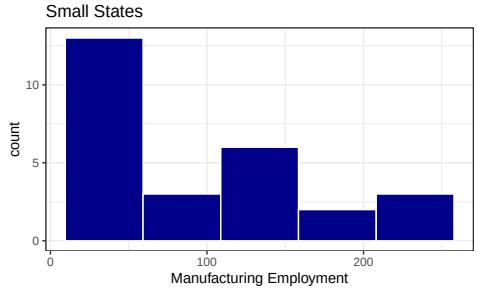
manufacturing_thousands and pop_category

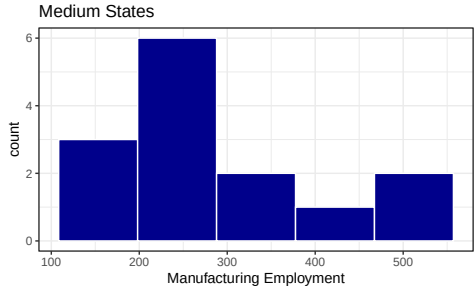
Bivariate Viz: Numerical x Categorical

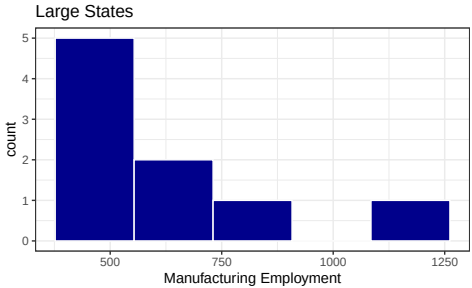
1. Using Facets to Extend Univariate Visualizations

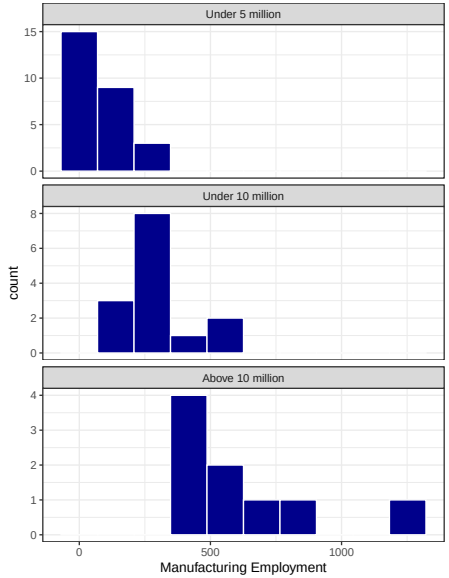
Make three histograms of manufacturing employment (5 bins)

1. pop_category = "Under 5 million"
2. pop_category = "Under 10 million"
3. pop_category = "Above 10 million"


Small States
count
10
5
0
0
100
200
Manufacturing Employment


Medium States
count
6
4
2
0
100
200
300
400
500
Manufacturing Employment


Large States
count
5
4
3
2
1
0
500
750
1000
1250
Manufacturing Employment


Under 5 million
Under 10 million
Above 10 million
count
Manufacturing Employment
0
500
1000

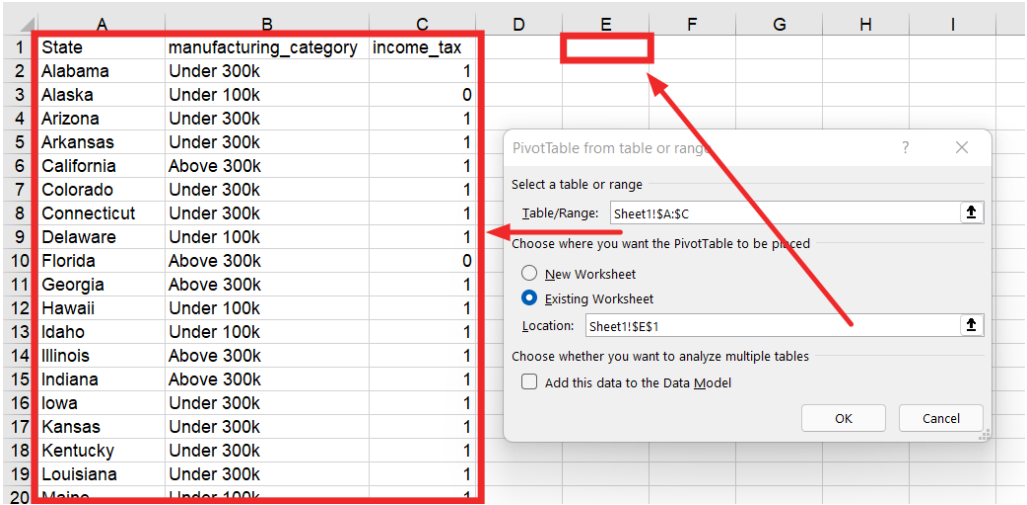
Bivariate Viz: Categorical x Categorical

1. Using Facets to Extend Univariate Visualizations

Make two bar plots of manufacturing category

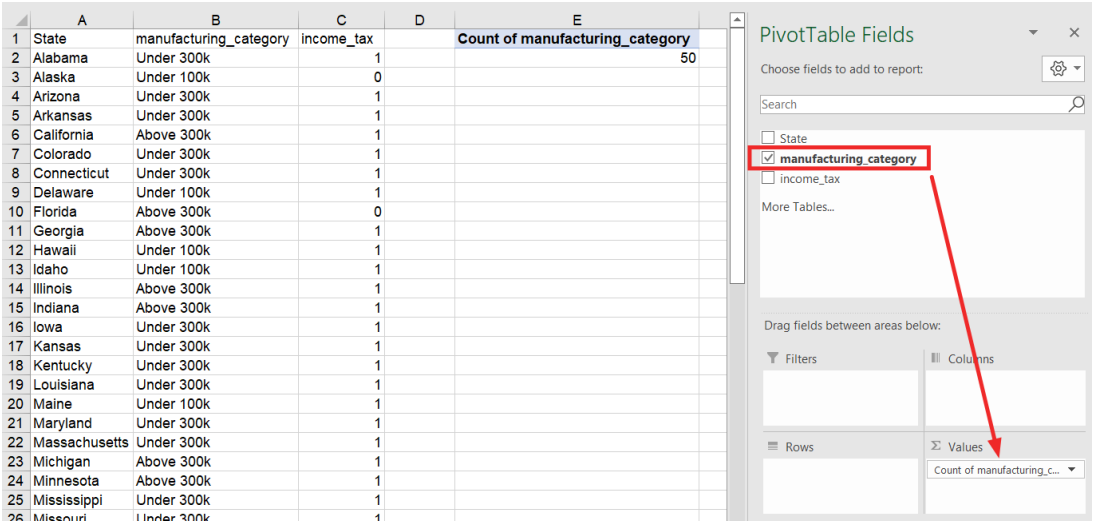
1. One for states with an income tax, and
2. One for states without an income tax.

1. Copy the data, insert a pivot table



	A	B	C
1	State	manufacturing_category	income_tax
2	Alabama	Under 300k	1
3	Alaska	Under 100k	0
4	Arizona	Under 300k	1
5	Arkansas	Under 300k	1
6	California	Above 300k	1
7	Colorado	Under 300k	1
8	Connecticut	Under 300k	1
9	Delaware	Under 100k	1
10	Florida	Above 300k	0
11	Georgia	Above 300k	1
12	Hawaii	Under 100k	1
13	Idaho	Under 100k	1
14	Illinois	Above 300k	1
15	Indiana	Above 300k	1
16	Iowa	Under 300k	1
17	Kansas	Under 300k	1
18	Kentucky	Under 300k	1
19	Louisiana	Under 300k	1
20	Maine	Under 100k	1

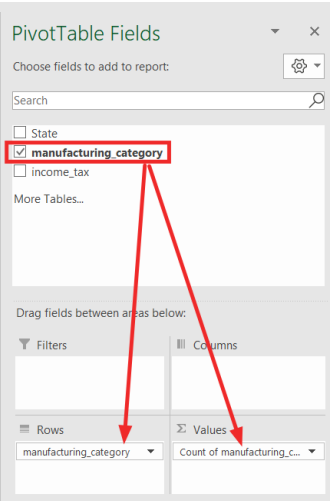
2. Choose the variable to count



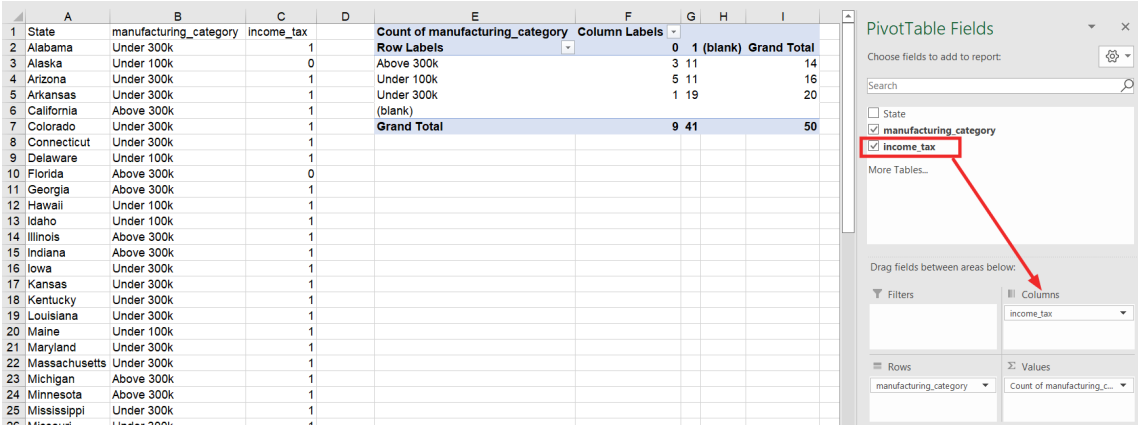
	A	B	C	D	E
1	State	manufacturing_category	income_tax		Count of manufacturing_category
2	Alabama	Under 300k	1		50
3	Alaska	Under 100k	0		
4	Arizona	Under 300k	1		
5	Arkansas	Under 300k	1		
6	California	Above 300k	1		
7	Colorado	Under 300k	1		
8	Connecticut	Under 300k	1		
9	Delaware	Under 100k	1		
10	Florida	Above 300k	0		
11	Georgia	Above 300k	1		
12	Hawaii	Under 100k	1		
13	Idaho	Under 100k	1		
14	Illinois	Above 300k	1		
15	Indiana	Above 300k	1		
16	Iowa	Under 300k	1		
17	Kansas	Under 300k	1		
18	Kentucky	Under 300k	1		
19	Louisiana	Under 300k	1		
20	Maine	Under 100k	1		
21	Maryland	Under 300k	1		
22	Massachusetts	Under 300k	1		
23	Michigan	Above 300k	1		
24	Minnesota	Above 300k	1		
25	Mississippi	Under 300k	1		
26	Missouri	Under 300k	1		

3. Count the levels separately

	A	B	C	D	E	F
1	State	manufacturing_category	income_tax		Row Labels	Count of manufacturing_category
2	Alabama	Under 300k	1		Above 300k	14
3	Alaska	Under 100k	0		Under 100k	16
4	Arizona	Under 300k	1		Under 300k	20
5	Arkansas	Under 300k	1		(blank)	
6	California	Above 300k	1		Grand Total	50
7	Colorado	Under 300k	1			
8	Connecticut	Under 300k	1			
9	Delaware	Under 100k	1			
10	Florida	Above 300k	0			
11	Georgia	Above 300k	1			
12	Hawaii	Under 100k	1			
13	Idaho	Under 100k	1			
14	Illinois	Above 300k	1			
15	Indiana	Above 300k	1			
16	Iowa	Under 300k	1			
17	Kansas	Under 300k	1			
18	Kentucky	Under 300k	1			
19	Louisiana	Under 300k	1			
20	Maine	Under 100k	1			
21	Maryland	Under 300k	1			
22	Massachusetts	Under 300k	1			
23	Michigan	Above 300k	1			
24	Minnesota	Above 300k	1			
25	Mississippi	Under 300k	1			
26	Missouri	Under 300k	1			

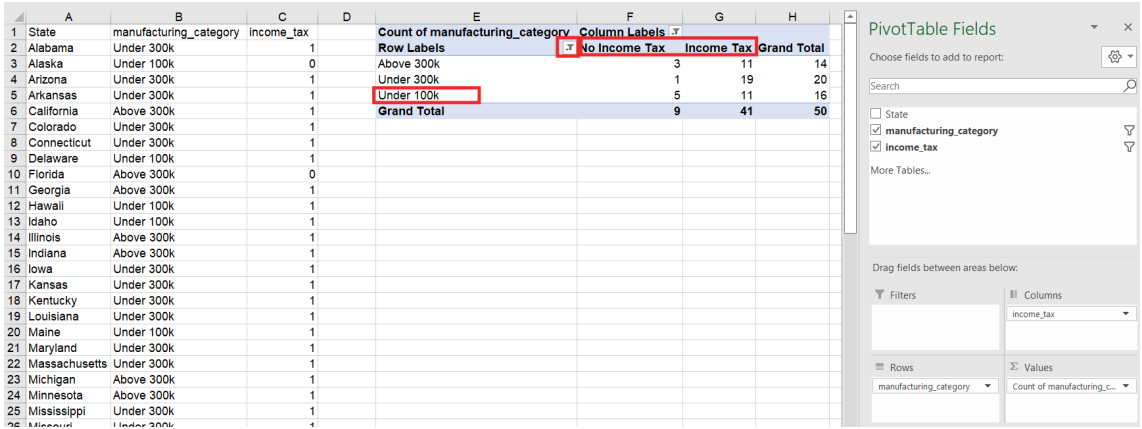


4. Add the second variable as columns



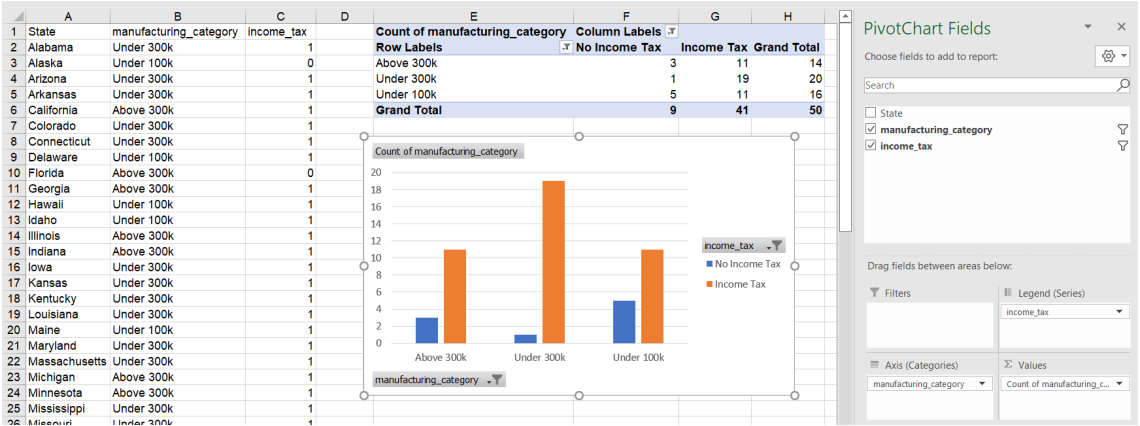
	A	B	C	D	E	F	G	H	I
1	State	manufacturing_category	income_tax		Count of manufacturing_category	Column Labels			
2	Alabama	Under 300k	1		Row Labels	0	1	(blank)	Grand Total
3	Alaska	Under 100k	0		Above 300k	3	11		14
4	Arizona	Under 300k	1		Under 100k	5	11		16
5	Arkansas	Under 300k	1		Under 300k	1	19		20
6	California	Above 300k	1		(blank)				
7	Colorado	Under 300k	1		Grand Total	9	41		50
8	Connecticut	Under 300k	1						
9	Delaware	Under 100k	1						
10	Florida	Above 300k	0						
11	Georgia	Above 300k	1						
12	Hawaii	Under 100k	1						
13	Idaho	Under 100k	1						
14	Illinois	Above 300k	1						
15	Indiana	Above 300k	1						
16	Iowa	Under 300k	1						
17	Kansas	Under 300k	1						
18	Kentucky	Under 300k	1						
19	Louisiana	Under 300k	1						
20	Maine	Under 100k	1						
21	Maryland	Under 300k	1						
22	Massachusetts	Under 300k	1						
23	Michigan	Above 300k	1						
24	Minnesota	Above 300k	1						
25	Mississippi	Under 300k	1						

5. Clean up the row order and labels



	A	B	C	D	E	F	G	H
1	State	manufacturing_category	income_tax		Count of manufacturing_category	Column Labels		
2	Alabama	Under 300k	1		Row Labels	No Income Tax	Income Tax	Grand Total
3	Alaska	Under 100k	0		Above 300k	3	11	14
4	Arizona	Under 300k	1		Under 300k	1	19	20
5	Arkansas	Under 300k	1		Under 100k	5	11	16
6	California	Above 300k	1		Grand Total	9	41	50
7	Colorado	Under 300k	1					
8	Connecticut	Under 300k	1					
9	Delaware	Under 100k	1					
10	Florida	Above 300k	0					
11	Georgia	Above 300k	1					
12	Hawaii	Under 100k	1					
13	Idaho	Under 100k	1					
14	Illinois	Above 300k	1					
15	Indiana	Above 300k	1					
16	Iowa	Under 300k	1					
17	Kansas	Under 300k	1					
18	Kentucky	Under 300k	1					
19	Louisiana	Under 300k	1					
20	Maine	Under 100k	1					
21	Maryland	Under 300k	1					
22	Massachusetts	Under 300k	1					
23	Michigan	Above 300k	1					
24	Minnesota	Above 300k	1					
25	Mississippi	Under 300k	1					
26	Missouri	Under 300k	1					

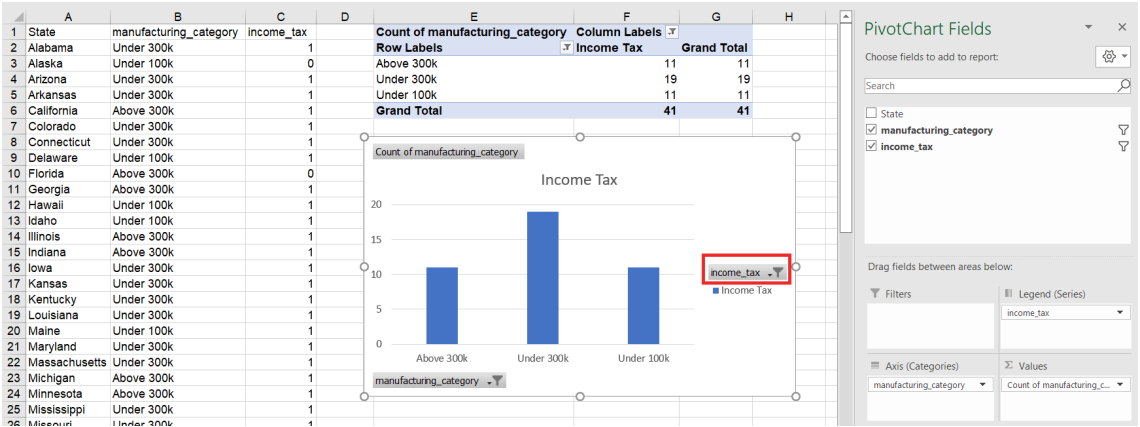
6. Insert a bar plot



	A	B	C
1	State	manufacturing_category	income_tax
2	Alabama	Under 300k	1
3	Alaska	Under 100k	0
4	Arizona	Under 300k	1
5	Arkansas	Under 300k	1
6	California	Above 300k	1
7	Colorado	Under 300k	1
8	Connecticut	Under 300k	1
9	Delaware	Under 100k	1
10	Florida	Above 300k	0
11	Georgia	Above 300k	1
12	Hawaii	Under 100k	1
13	Idaho	Under 100k	1
14	Illinois	Above 300k	1
15	Indiana	Above 300k	1
16	Iowa	Under 300k	1
17	Kansas	Under 300k	1
18	Kentucky	Under 300k	1
19	Louisiana	Under 300k	1
20	Maine	Under 100k	1
21	Maryland	Under 300k	1
22	Massachusetts	Under 300k	1
23	Michigan	Above 300k	1
24	Minnesota	Above 300k	1
25	Mississippi	Under 300k	1
26	Missouri	Under 300k	1

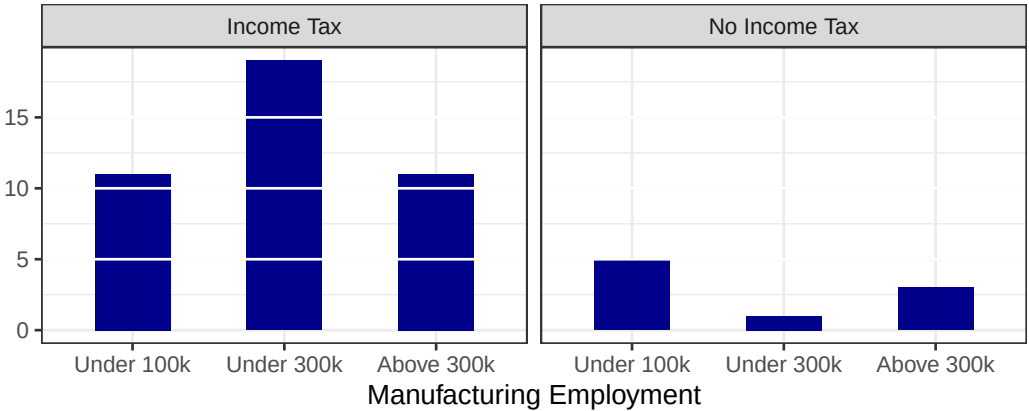
Count of manufacturing_category	Column Labels		
Row Labels	No Income Tax	Income Tax	Grand Total
Above 300k	3	11	14
Under 300k	1	19	20
Under 100k	5	11	16
Grand Total	9	41	50

7. Use the chart filter to separate plots



	A	B	C	D	E	F	G	H
1	State	manufacturing_category	income_tax		Count of manufacturing_category	Column Labels		
2	Alabama	Under 300k	1		Row Labels	Income Tax	Grand Total	
3	Alaska	Under 100k	0		Above 300k	11	11	
4	Arizona	Under 300k	1		Under 300k	19	19	
5	Arkansas	Under 300k	1		Under 100k	11	11	
6	California	Above 300k	1		Grand Total	41	41	
7	Colorado	Under 300k	1					
8	Connecticut	Under 300k	1					
9	Delaware	Under 100k	1					
10	Florida	Above 300k	0					
11	Georgia	Above 300k	1					
12	Hawaii	Under 100k	1					
13	Idaho	Under 100k	1					
14	Illinois	Above 300k	1					
15	Indiana	Above 300k	1					
16	Iowa	Under 300k	1					
17	Kansas	Under 300k	1					
18	Kentucky	Under 300k	1					
19	Louisiana	Under 300k	1					
20	Maine	Under 100k	1					
21	Maryland	Under 300k	1					
22	Massachusetts	Under 300k	1					
23	Michigan	Above 300k	1					
24	Minnesota	Above 300k	1					
25	Mississippi	Under 300k	1					
26	Missouri	Under 300k	1					

Bivariate Viz: Categorical x Categorical



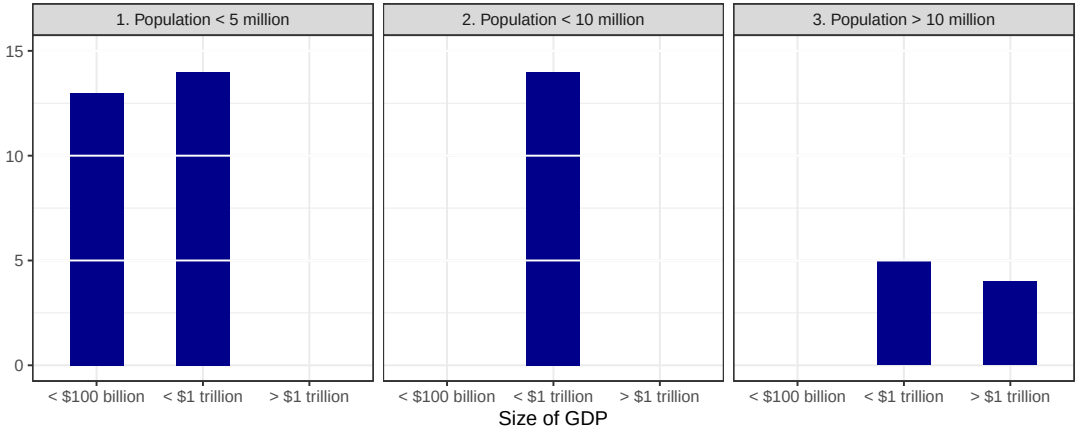
Bivariate Viz: Categorical x Categorical

1. Using Facets to Extend Univariate Visualizations

Make three bar plots of GDP category

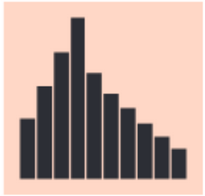
1. pop_category = "Under 5 million"
2. pop_category = "Under 10 million"
3. pop_category = "Above 10 million"

Bivariate Viz: Categorical x Categorical



1. Facets

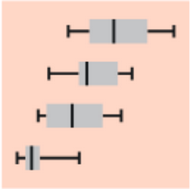
Histogram



Ordered column

2.

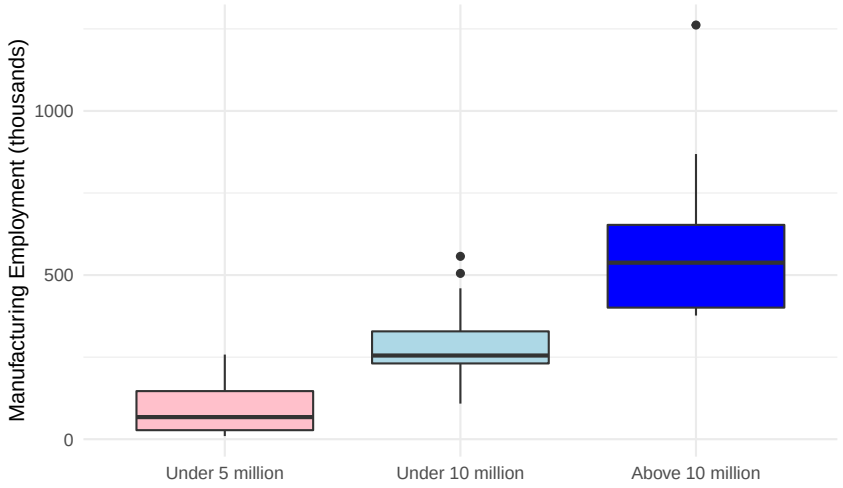
Boxplot

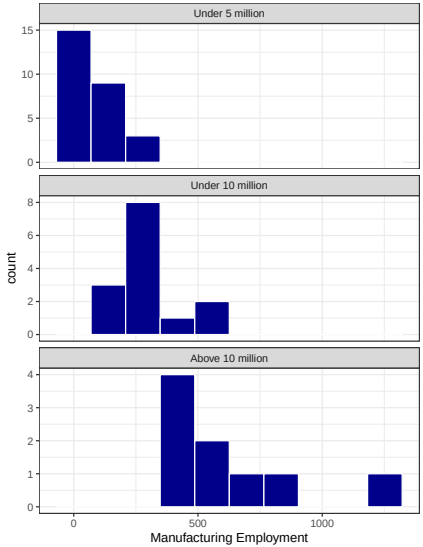


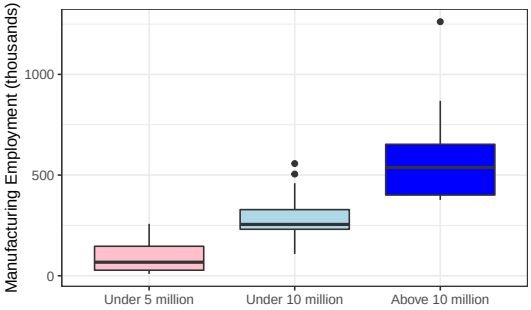
3.

Scatterplot

Bivariate Viz: Numerical x Categorical

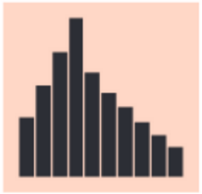



Under 5 million
Under 10 million
Above 10 million
count
Manufacturing Employment
0
500
1000
15
10
5
8
6
4
2
4
3
2
1


Manufacturing Employment (thousands)
1000
500
0
Under 5 million
Under 10 million
Above 10 million

1. Facets

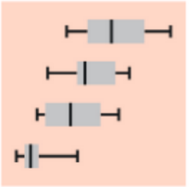
Histogram



Ordered column

2.

Boxplot



3.

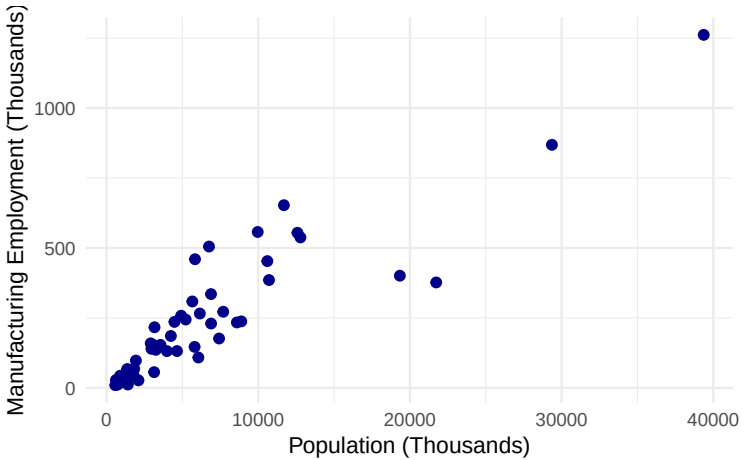
Scatterplot

Using Scatter Plots: Numerical x Numerical

Do states with bigger populations have higher levels of employment in manufacturing?

Make a scatter plot of manufacturing employment and population.

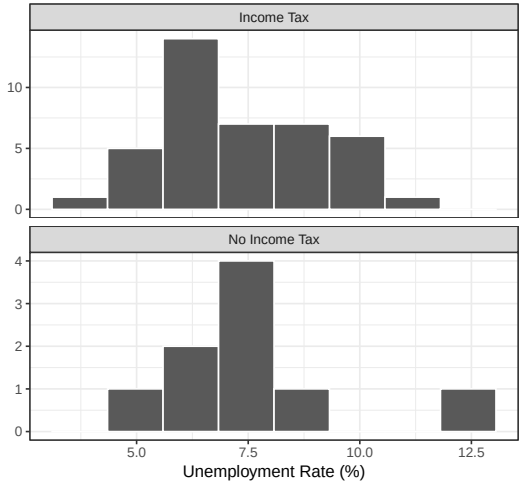
Bivariate Viz: Numerical x Numerical

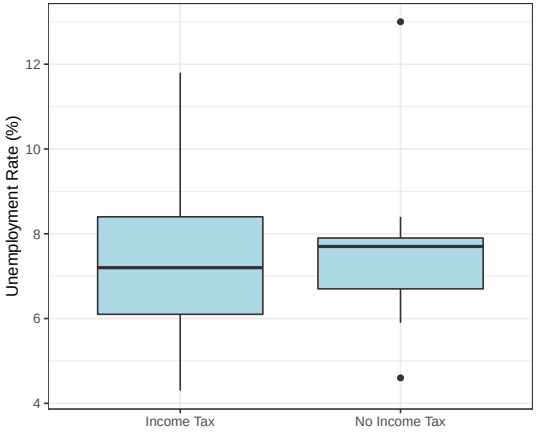


Time to Practice!

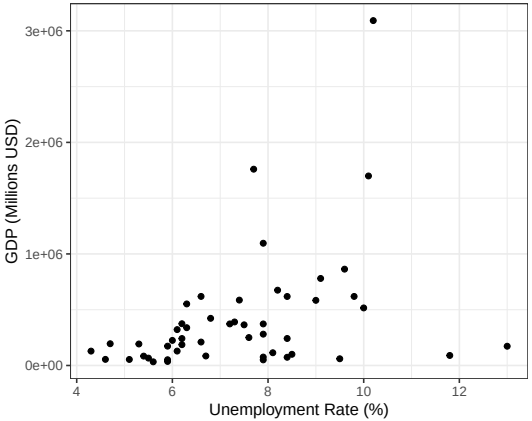
1. Make two histograms of unemployment, one for states with an income tax and one for states without.
2. Remake the above as a box plot
3. Make and analyze the following four scatter plots:
    - GDP (actual) x Unemployment
    - GDP (actual) x Bachelors' Degrees
    - GDP (rate) x Unemployment
    - GDP (rate) x Bachelors' Degrees

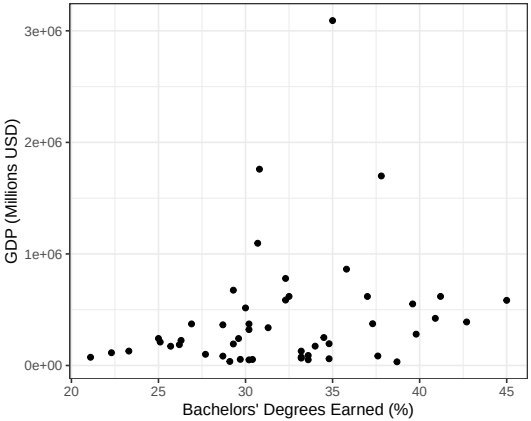
Unemployment x Income Taxes





Scatter plots of GDP (actual)





Scatter plots of GDP (rate)