Political Science 30: Political Inquiry
Section 7

Taylor Carlson
tncarlson@ucsd.edu

"Errors using inadequate data are much less than those using no data at all." — Charles Babbage

Learning Outcomes

By the end of section today, you should:

- Be able to conduct and interpret a chi squared test
- Be able to communicate data effectively using graphs

Warm Up

For each variable listed below, list (a) the **type of variable** it is, (b) a measure of central tendency for that type of variable, and (c) a measure of dispersion for that type of variable.

- Marital status (married, divorced, widowed, single, domestic partnership)
- Age (years: 0, 1, 2, 3, 4...)
- Education (no formal education, grade school only, high school, some college, bachelors, graduate or professional degree)
- Voter turnout (percentage of the voting eligible population that voted)
- Vote choice (Trump, Clinton)
- Ideology (Extremely liberal, liberal, moderate, conservative, extremely conservative)

Announcements

- HW2 passed back at the end of section today
- **HW3 is due on Wednesday!**
  - Make sure that you have an *interval* DV!!
  - See the graphing practice exercise online for plenty of help with the data portion
  - See examples from section, exercises, and class
  - Extra OH Monday 3:30-4:30

Graphing Data

- Bar Graphs

Graphing Data

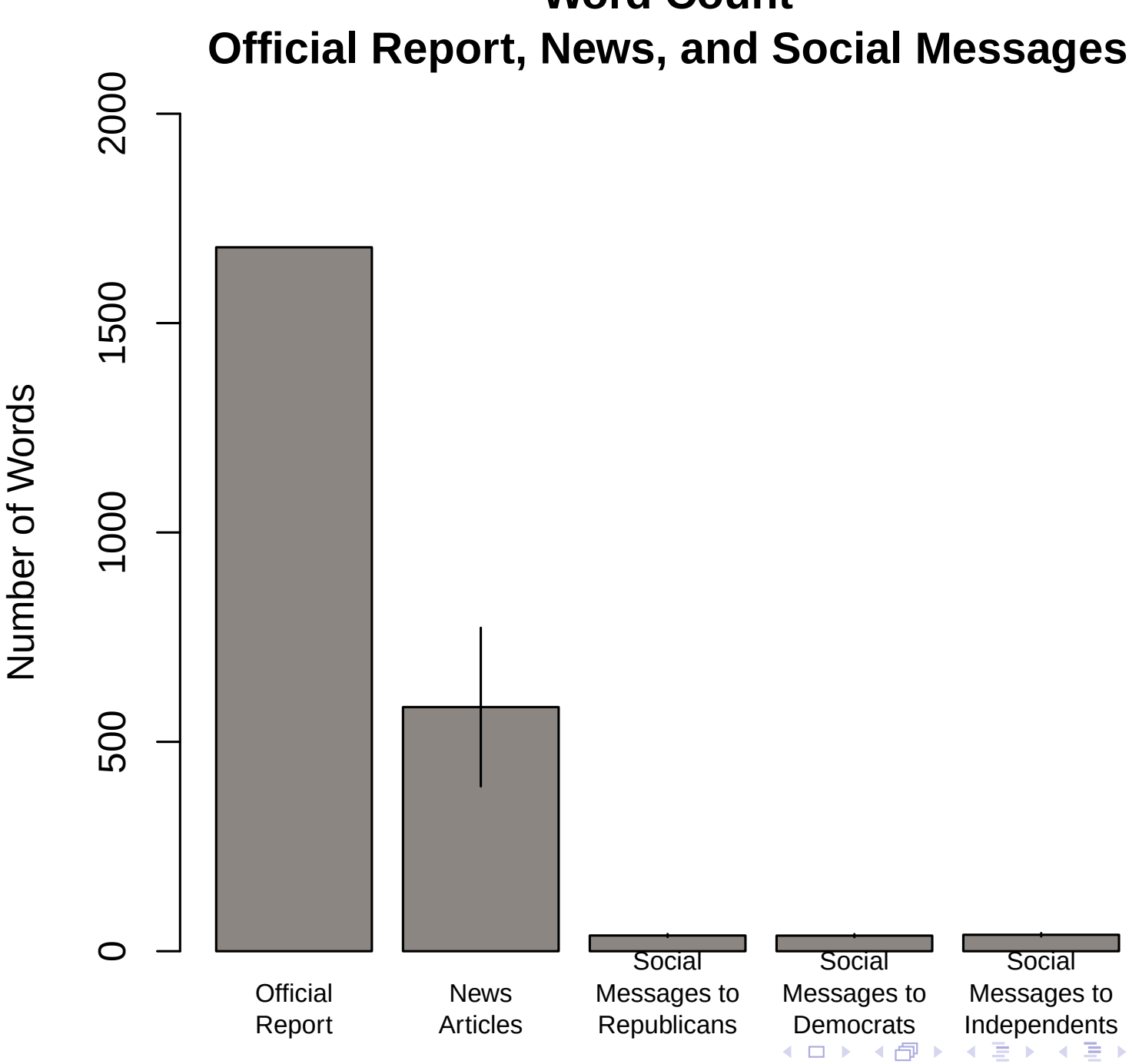
- Bar Graphs
  - Interval DV, nominal or ordinal IV
  - Stata Code: `graph bar (mean) DV, over(IV)`
  - Show the mean of your DV for each value of your IV

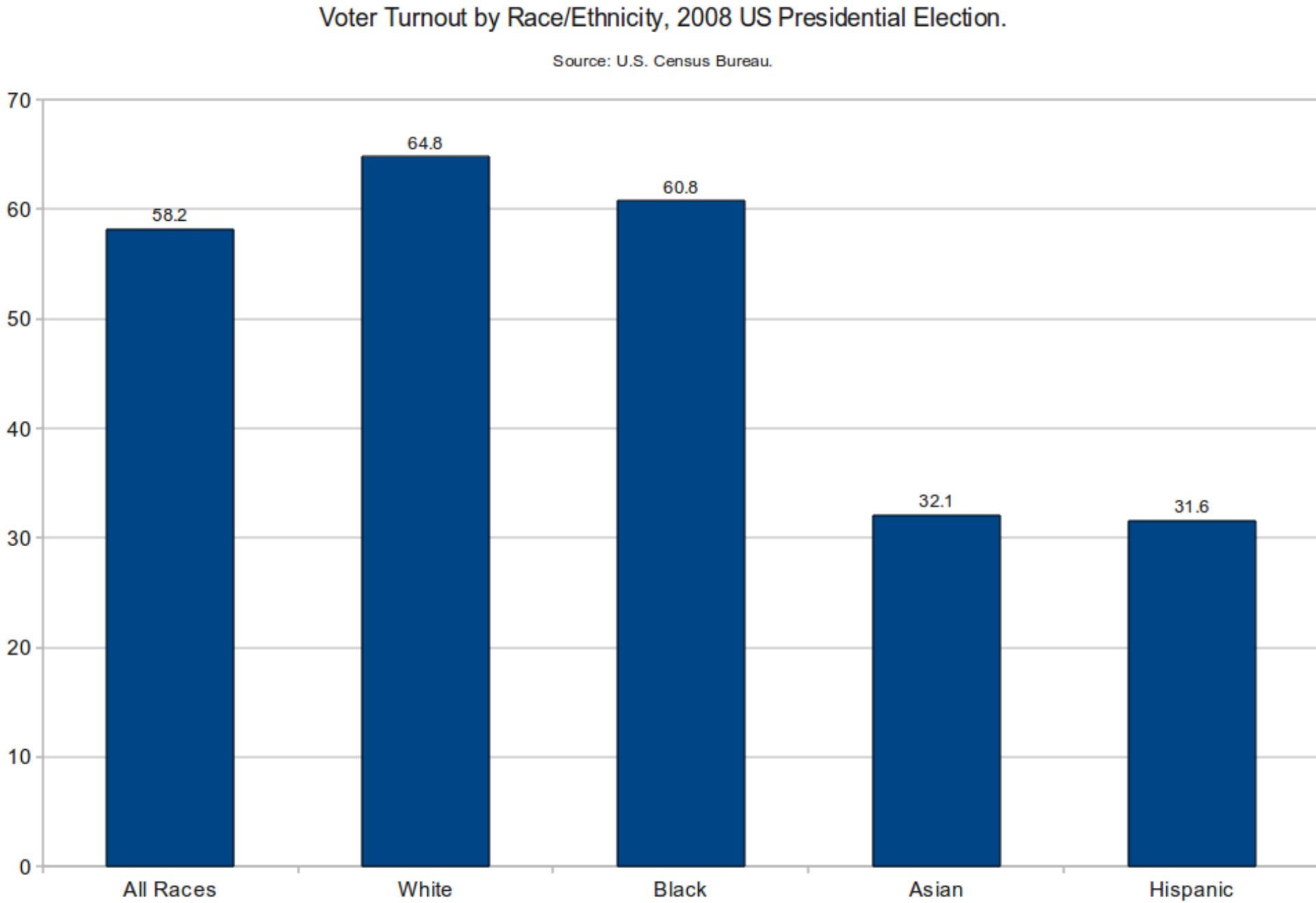
Graphing Data

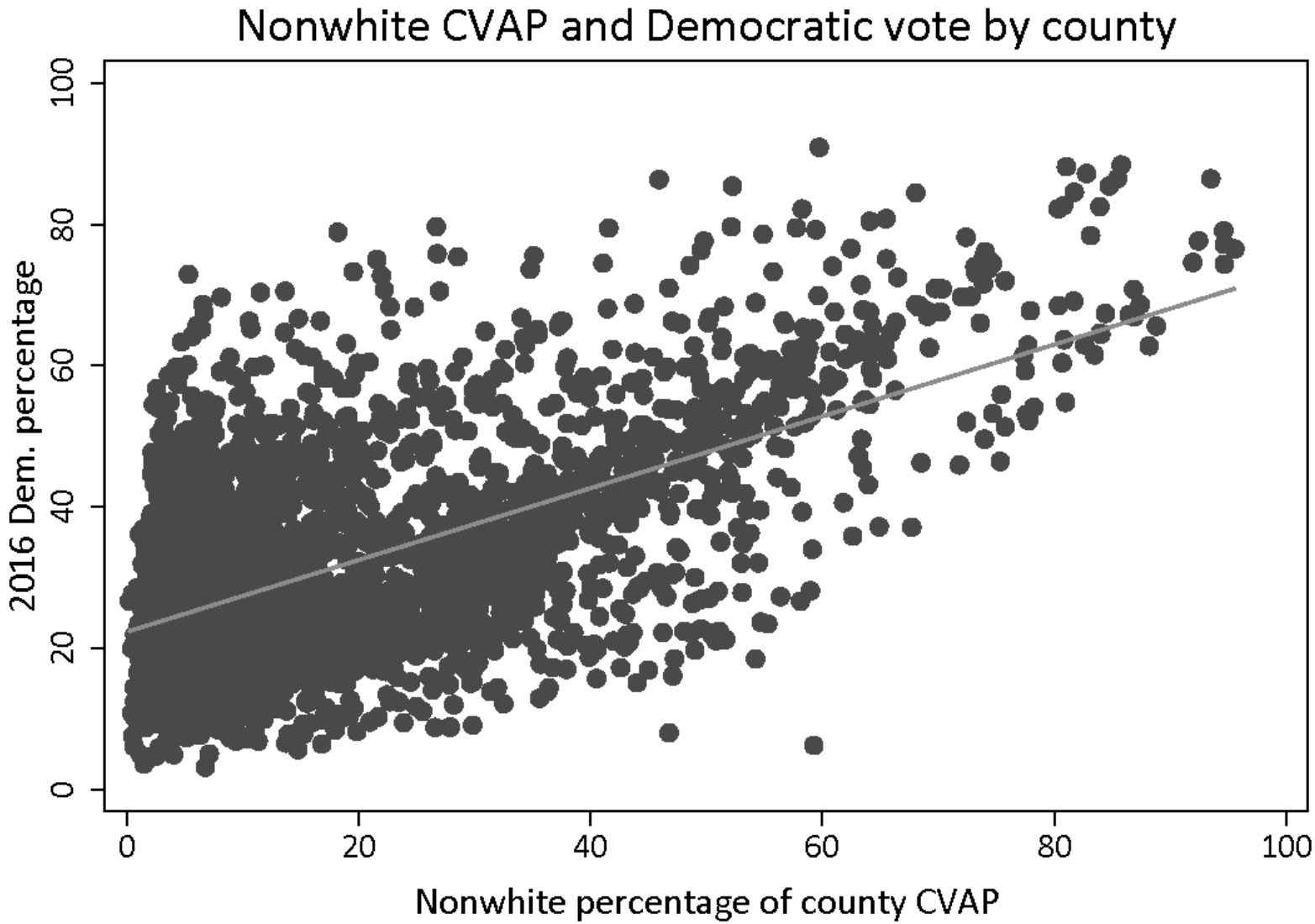
- Bar Graphs
  - Interval DV, nominal or ordinal IV
  - Stata Code: `graph bar (mean) DV, over(IV)`
  - Show the mean of your DV for each value of your IV
- Scatter plots

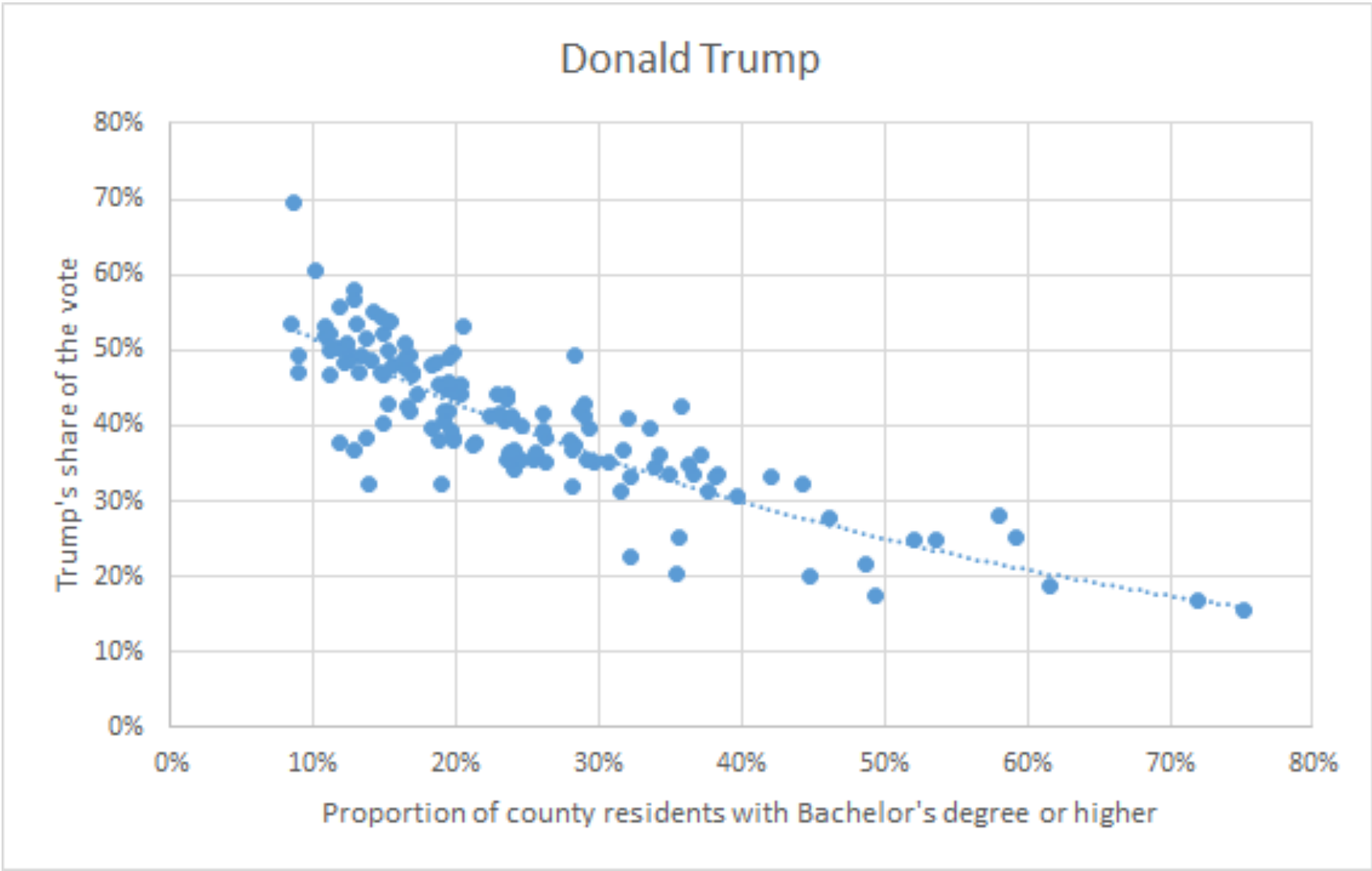
Graphing Data

- Bar Graphs
  - Interval DV, nominal or ordinal IV
  - Stata Code: `graph bar (mean) DV, over(IV)`
  - Show the mean of your DV for each value of your IV
- Scatter plots
  - Interval DV, interval IV
  - Stata code: `scatter DV IV`
  - DV goes on the Y-axis, IV goes on the X-axis
  - Each point is an observation at its level of the IV and DV


Word Count
Official Report, News, and Social Messages
Number of Words
0
500
1000
1500
2000
Official Report
News Articles
Social Messages to Republicans
Social Messages to Democrats
Social Messages to Independents


Voter Turnout by Race/Ethnicity, 2008 US Presidential Election.
Source: U.S. Census Bureau.
70
60
50
40
30
20
10
0
58.2
64.8
60.8
32.1
31.6
All Races
White
Black
Asian
Hispanic


Nonwhite CVAP and Democratic vote by county
2016 Dem. percentage
0
20
40
60
80
100
Nonwhite percentage of county CVAP


Donald Trump
Trump's share of the vote
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
Proportion of county residents with Bachelor's degree or higher

With the person next to you...

- Discuss your hypothesis for your data project
- Explain which variables you are using and come up with the types of variables
- Figure out which type of graph you'll need to use to visualize the relationship between your variables

Questions?

Chi Square Test

When we want to explore the relationship between a nominal IV and a nominal DV, we have two options:

- Difference of proportions (2 groups, nominal DV)
- Chi square test (2+ groups, nominal DV)

Chi Square Test

	IV Value 1	IV Value 2	IV Value 3	Total
DV Value 1				
DV Value 2				
Total				N=

Chi Square Test

We begin just like all of our other tests — with our hypotheses:

Chi Square Test

We begin just like all of our other tests — with our hypotheses:

- H_0: All categories of the IV are proportionally equal. The difference in proportions between categories is equal to 0.
  - The IV and DV are independent.

Chi Square Test

We begin just like all of our other tests — with our hypotheses:

- H_0: All categories of the IV are proportionally equal. The difference in proportions between categories is equal to 0.
  - The IV and DV are independent.
- H_1: All categories of the DV are not proportionally equal. The difference in proportions between categories is not equal to 0.
  - The IV and DV are not independent

Chi Square Test

$$\chi^2 = \sum_{k=1}^{k} \frac{(f_O - f_e)^2}{f_e}$$

- K = number of cells in the table
- f_o=frequencies/counts **observed** in each category
- f_e=frequencies/counts **expected** in each category if our null hypothesis is right: that the IV and DV are independent (not related)

Chi Square Test: Finding the Pieces

We get to **observe** f_O because that is what appears in our table — it's what we actually see in the data.

Chi Square Test: Finding the Pieces

We get to **observe** f_O because that is what appears in our table — it's what we actually see in the data.

But what about the **expected** counts (f_e)?

$$\frac{RowTotalxColumnTotal}{TotalN}$$

Chi Square Test: Overview of the Steps

- Write null and alternative hypotheses
- Calculate χ^2
  - Calculate expected counts (f_e)
  - Subtract each one from its observed count (f_O)
  - Square this difference
  - Divide by the expected count (f_e)
  - Sum all of these quotients together
- Calculate the degrees of freedom (Rows-1) x (Columns -1)
- Find the threshold value for your degrees of freedom in the Chi Square Table
- Compare your X^2 to the threshold value
  - If $\chi^2 > threshold$, Reject the null hypothesis
  - If $\chi^2 < threshold$, fail to reject null hypothesis

Chi Square Test: Example

H_0: Community type (IV) is independent of presidential vote in 2016 (DV).
H_1: Community type (IV) affects presidential vote in 2016.

Chi Square Test: Example

H_0: Community type (IV) is independent of presidential vote in 2016 (DV).
H_1: Community type (IV) affects presidential vote in 2016.

Here are all of our **observed counts**: f_O

	Urban	Suburban	Rural	Total
Clinton	46	37	20	103
Trump	71	83	88	242
Total	117	120	108	N=345

Chi Square Test: Example

H_0: Community type (IV) is independent of presidential vote in 2016 (DV).
H_1: Community type (IV) affects presidential vote in 2016.

Here are all of our **observed counts**: f_O

	Urban	Suburban	Rural	Total
Clinton	46	37	20	103
Trump	71	83	88	242
Total	117	120	108	N=345

Our next step is going to be to calculate the **expected counts** (f_e). Let's make a new table with what we would expect if there was no relationship.

Chi Square Test: Example — Expected Counts f_e

$$\frac{RowTotalxColumnTotal}{TotalN}$$

Chi Square Test: Example — Expected Counts f_e

$$\frac{RowTotalxColumnTotal}{TotalN}$$

For Each Cell:

	Urban	Suburban	Rural	Total
Clinton				103
Trump				242
Total	117	120	108	N=345

Chi Square Test: Expected Counts f_e

$$\frac{RowTotalxColumnTotal}{TotalN}$$

For Each Cell:

	Urban	Suburban	Rural	Total
Clinton	(103x117)/345	(103x120)/345	(103x108)/345	103
Trump				242
Total	117	120	108	N=345

Chi Square Test: Expected Counts f_e

$$\frac{RowTotalxColumnTotal}{TotalN}$$

For Each Cell:

	Urban	Suburban	Rural	Total
Clinton	(103x117)/345	(103x120)/345	(103x108)/345	103
Trump	(242x117)/345	(242x120)/345	(242x108)/345	242
Total	117	120	108	N=345

Chi Square Test: Expected Counts f_e

$$\frac{RowTotalxColumnTotal}{TotalN}$$

	Urban	Suburban	Rural	Total
Clinton	34.93	35.83	32.24	103
Trump	82.07	84.17	75.76	242
Total	117	120	108	N=345

Chi Square Test: f_O - f_e

Subtract each f_e from the observed f_O from the original table:

	Urban	Suburban	Rural	Total
Clinton	46-34.93	37-35.83	20-32.24	103
Trump	71-82.07	83-84.17	88-75.76	242
Total	117	120	108	N=345

Chi Square Test: f_O - f_e

Subtract each f_e from the observed f_O from the original table:

	Urban	Suburban	Rural	Total
Clinton	11.07	1.17	-12.24	103
Trump	-11.07	-1.17	12.24	242
Total	117	120	108	N=345

Chi Square Test: $(f_O - f_e)^2$

Square each difference

	Urban	Suburban	Rural	Total
Clinton	122.54	1.37	149.8	103
Trump	122.54	1.37	149.8	242
Total	117	120	108	N=345

Chi Square Test: $(f_O - f_e)^2/f_e$

Divide the number in each cell by its expected count f_e

	Urban	Suburban	Rural	Total
Clinton	122.54/34.93	1.37/35.83	149.8/32.24	103
Trump	122.54/82.07	1.37/84.17	149.8/75.76	242
Total	117	120	108	N=345

Chi Square Test: $(f_O - f_e)^2/f_e$

Divide the number in each cell by its expected count f_e

	Urban	Suburban	Rural	Total
Clinton	3.51	0.04	4.65	103
Trump	1.49	0.02	1.98	242
Total	117	120	108	N=345

Chi Square Test: $\Sigma(f_O - f_e)^2/f_e$

Add up all of the cells (NOT the totals)

	Urban	Suburban	Rural	Total
Clinton	3.51	0.04	4.65	103
Trump	1.49	0.02	1.98	242
Total	117	120	108	N=345

$3.51 + 1.49 + 0.04 + 0.02 + 4.65 + 1.98 =$

Chi Square Test: $\Sigma(f_O - f_e)^2/f_e$

Add up all of the cells (NOT the totals)

	Urban	Suburban	Rural	Total
Clinton	3.51	0.04	4.65	103
Trump	1.49	0.02	1.98	242
Total	117	120	108	N=345

$3.51 + 1.49 + 0.04 + 0.02 + 4.65 + 1.98 = \mathbf{11.69}$

Chi Square Test: $\Sigma(f_O - f_e)^2/f_e$

Add up all of the cells (NOT the totals)

	Urban	Suburban	Rural	Total
Clinton	3.51	0.04	4.65	103
Trump	1.49	0.02	1.98	242
Total	117	120	108	N=345

$$3.51 + 1.49 + 0.04 + 0.02 + 4.65 + 1.98 = \mathbf{11.69}$$

$$\chi^2 = 11.69$$

Chi Square Test: Degrees of Freedom

Now figure out how many degrees of freedom you have:

$$(\text{Rows-1}) \times (\text{Columns-1})$$

Chi Square Test: Degrees of Freedom

Now figure out how many degrees of freedom you have:

$$\text{(Rows-1) x (Columns-1)}$$

$$\text{(2-1) x (3-1)} = 1 \text{ x } 2 = 2$$

Chi Square Test: Degrees of Freedom

Now figure out how many degrees of freedom you have:

$$\text{(Rows-1) x (Columns-1)}$$

$$\text{(2-1) x (3-1)} = 1 \text{ x } 2 = 2$$

Chi Square Test: Find Threshold Value

Now look up your threshold value in the Chi Square Table based on the degrees of freedom (2 in our case):

Degrees of Freedom	Threshold Value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92

Chi Square Test: Compare χ^2 to Threshold

Compare our χ^2 to our threshold. $\chi^2 = 11.69$

Degrees of Freedom	Threshold Value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92

Chi Square Test: Compare χ^2 to Threshold

Compare our χ^2 to our threshold. $\chi^2 = 11.69$

Degrees of Freedom	Threshold Value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92

Is $\chi^2 >$ *threshold*?

Chi Square Test: Compare χ^2 to Threshold

Compare our χ^2 to our threshold. $\chi^2 = 11.69$

Degrees of Freedom	Threshold Value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92

Is $\chi^2 > \textit{threshold}$? $11.69 > 5.99$

Chi Square Test: Compare χ^2 to Threshold

Compare our χ^2 to our threshold. $\chi^2 = 11.69$

Degrees of Freedom	Threshold Value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92

Is $\chi^2 > \textit{threshold}$? $11.69 > 5.99$ Yes. Therefore, we can reject the null hypothesis and conclude with 95% confidence that presidential vote choice is not independent of community type.