

# Poli 30D Political Inquiry

## Descriptive Statistics & Visualization

Shane Xinyang Xuan  
ShaneXuan.com

October 15, 2016

## Contact Information

Shane Xinyang Xuan  
xxuan@ucsd.edu

We have someone to help you every day!

Professor Desposato	M	1330-1500 (Latin American Center)
Shane Xuan	Tu	1600-1800 (SSB332)
Cameron Sells	W	1000-1200 (SSB352)
Kelly Matush	Th	1500-1700 (SSB343)
Julia Clark	F	1200-1400 (SSB326)

Supplemental Materials

Our class oriented

ShaneXuan.com

UCLA SPSS starter kit

[www.ats.ucla.edu/stat/spss/sk/modules\\_sk.htm](http://www.ats.ucla.edu/stat/spss/sk/modules_sk.htm)

Princeton data analysis

<http://dss.princeton.edu/training/>

## Variable

- Nominal (categorical)  
i.e. Hillary, Donald, Gary, Jill
- Ordinal (can rank)  
i.e. strongly agree > agree > neutral > disagree > strongly disagree
- Interval (different by how much?)  
i.e. grade in school, happiness index, election fraud index

## Ratio Variable

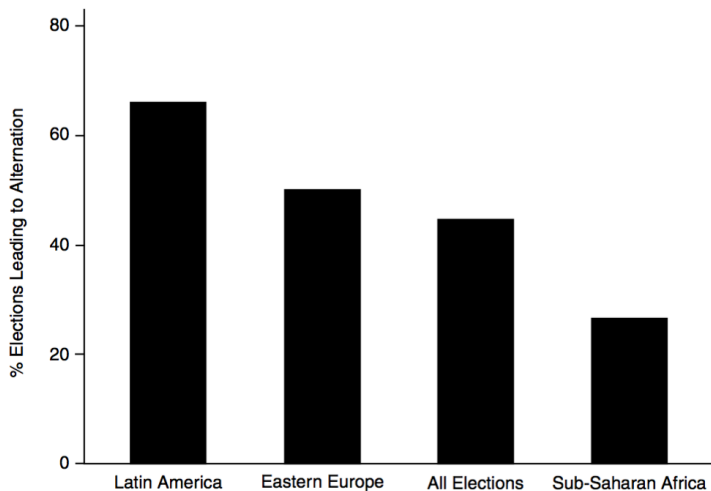
- Interval, but with a **meaningful** zero
- Does it make sense to **compare** this to 0?
- Could I **divide** this by a number or another variable, and would it still make sense?
- i.e. age, distance in miles

## Visualize our measurement

- Bar chart
- Scatterplot
- Pie chart
- Histogram
- Other visualization

Let's go through some examples!

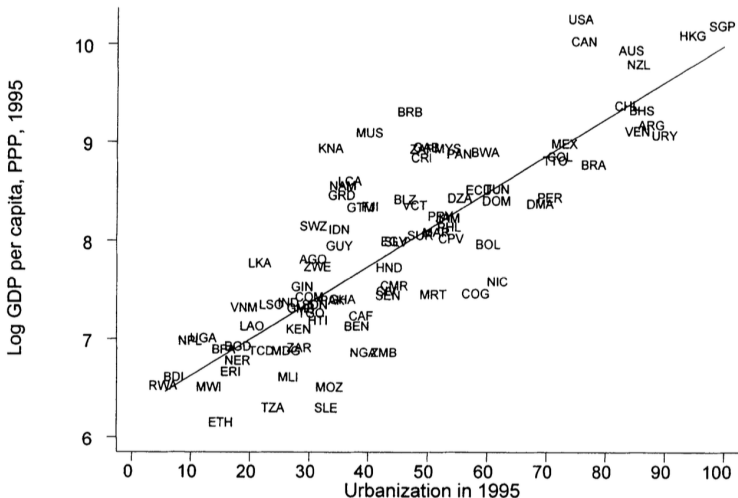
Explain the following bar chart to me (Arriola 2013)



# Scatter plot

Explain the following scatter plot to me

(Acemoglu, Johnson, Robinson 2002)



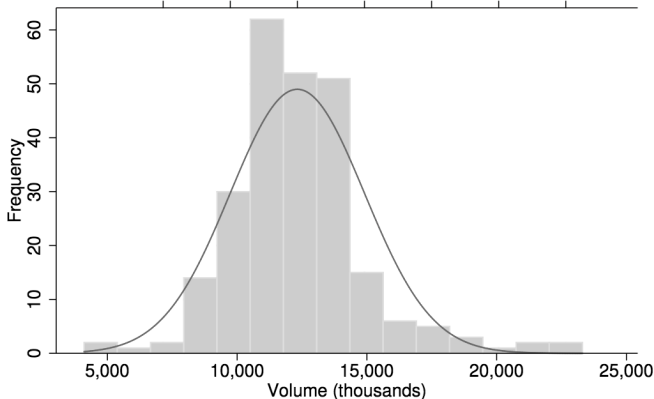
# Histogram

Explain the following histogram to me

(Yahoo! Finance and Commodity Systems)

S&P 500, January 2001 – December 2001

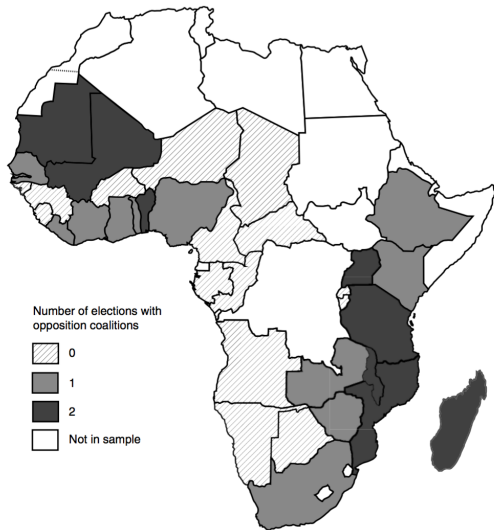
-2 s.d. -1 s.d. mean +1 s.d. +2 s.d. +3 s.d. +4 s.d.





# What else? Map!

Explain the following figure to me (Arriola 2013)



# Central Tendency

Now that we're done with visualization, let's dig into more concepts

- Mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Mode

In  $\{3, 4, 6, 6, 6, 7, 7, 9\}$ , the mode is 6 (most often)

- Median

When  $n = \text{odd}$ , check  $\frac{n+1}{2}$

When  $n = \text{even}$ , take the average of  $\frac{n}{2}$  and  $(\frac{n}{2} + 1)$

Positive Skew: Mean  $>$  Median

# Dispersion

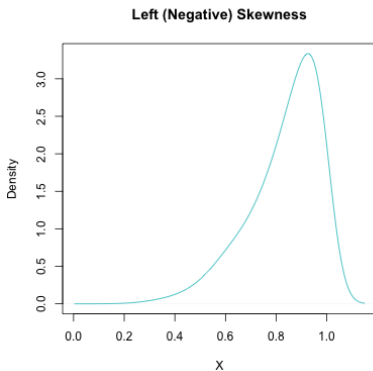
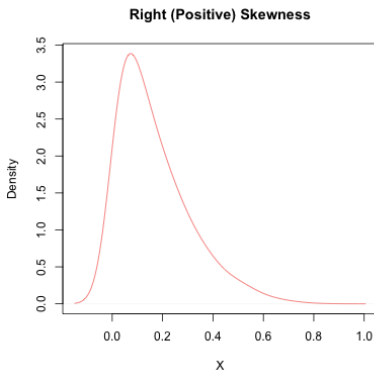
Positive Skew: Mean  $>$  Median

Negative Skew: Mean  $<$  Median

# Dispersion

**Positive Skew:** Mean  $>$  Median

**Negative Skew:** Mean  $<$  Median



# Variance & Standard Deviation

Variance is defined as

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}$$

Standard deviation is defined as

$$\begin{aligned}\sigma &\equiv \sqrt{\sigma^2} \\ &= \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}}\end{aligned}$$

# Variance & Standard Deviation

## Example

$x_i$	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
1		
2		
3		
4		
5		

Find the mean

$$\bar{X} = \frac{1 + 2 + 3 + 4 + 5}{5} = 3$$

# Variance & Standard Deviation

## Example

$x_i$	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
1	-2	
2	-1	
3	0	
4	1	
5	2	

Calculate the 2<sup>nd</sup> column

$$x_1 - \bar{X} = 1 - 3 = -2$$

$$x_2 - \bar{X} = 2 - 3 = -1$$

⋮

$$x_5 - \bar{X} = 5 - 3 = 2$$



# Variance & Standard Deviation

## Example

$x_i$	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
1	-2	4
2	-1	1
3	0	0
4	1	1
5	2	4

Square the 2<sup>nd</sup> column

$$(x_1 - \bar{X})^2 = (-2)^2 = 4$$

$$(x_2 - \bar{X})^2 = (-1)^2 = 1$$

⋮

$$(x_5 - \bar{X})^2 = 2^2 = 4$$

# Variance & Standard Deviation

## Example

$x_i$	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
1	-2	4
2	-1	1
3	0	0
4	1	1
5	2	4

Let me remind you the formula

$$\begin{aligned}\sigma^2 &= \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} \\ &= \frac{4 + 1 + 0 + 1 + 4}{5}\end{aligned}$$

I will see you next week!

Contact me if you have question  
xxuan@ucsd.edu