

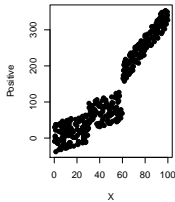
PLSC 308: Introduction to Political Research

Christopher Zorn

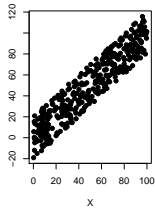
April 5, 2016

Types of Relationships

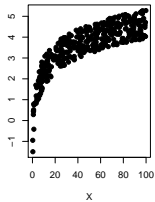
Monotonic



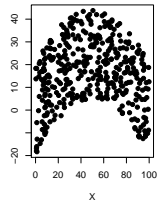
Linear



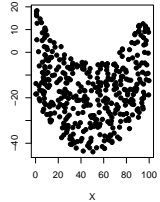
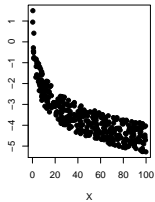
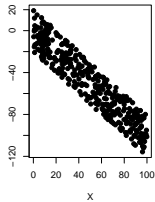
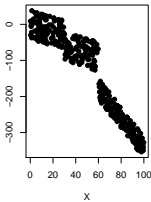
Non-Linear



Non-Monotonic

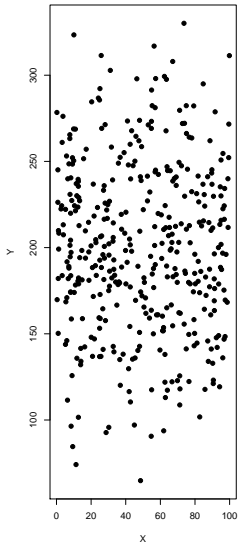


Negative

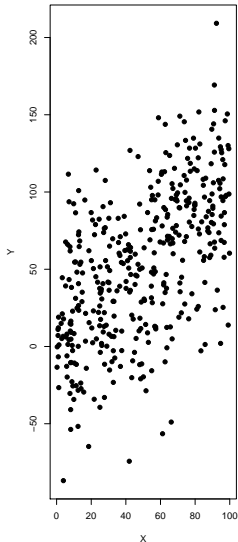


Strength of Relationships

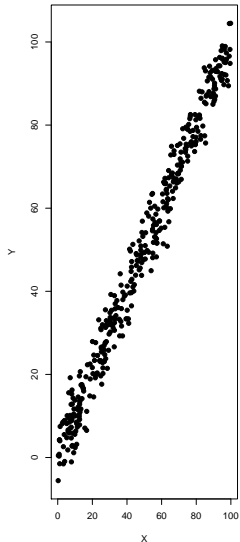
No Relationship



Weak Relationship



Strong Relationship



Methods of Assessing Association

- Tabular
- Statistical
- Graphical

Most common: Two-Way Frequency Tables (“Crosstabs”)

- Requires *nominal*- or *ordinal*-level data...
- Rows / columns denote categories (or intervals) of Y and X respectively
- Cell entries indicate frequencies of observations that meet both conditions...

Crosstabs: An Example

Civil War?	Sub-Saharan?		Total
	No	Yes	
No	5	25	30
(Row)	(0.17)	(0.83)	(1.00)
[Column]	[0.83]	[0.68]	[0.70]
{Cell}	{0.12}	{0.58}	{0.70}
Yes	1	12	13
(Row)	(0.08)	(0.92)	(1.00)
[Column]	[0.17]	[0.32]	[0.30]
{Cell}	{0.02}	{0.28}	{0.30}
Total	6	37	43
	(0.14)	(0.86)	(1.00)
	[1.00]	[1.00]	[1.00]
	{0.14}	{0.86}	{1.00}

Crosstabs: Example Code

```
> TABLE<-with(Africa, table(internalwar,subsaharan))  
> TABLE
```

	subsaharan	
internalwar	Not Sub-Saharan	Sub-Saharan
0	5	25
1	1	12

Row and Column Percentages

```
> prop.table(TABLE,1)*100
```

```
      subsaharan
```

internalwar	Not Sub-Saharan	Sub-Saharan
0	16.666667	83.333333
1	7.692308	92.307692

```
> prop.table(TABLE,2)*100
```

```
      subsaharan
```

internalwar	Not Sub-Saharan	Sub-Saharan
0	83.33333	67.56757
1	16.66667	32.43243

Crosstabs: General Guidelines

- Nominal categories = no indication of “direction”
- Ordinal categories should appear in order
- Continuous variables require “binning” ...
- Are related to statistics (e.g., χ^2 ; see below)

Statistical Measures of Association

The general idea:

- If two variables are unrelated, then we should see an “even” distribution of cases on each, irrespective of the values of the other.
- If we observe something other than such an “even” distribution, then the variables are not unrelated.
- Formally: $H_0 : f(Y|X) = f(Y)$

Example: Chi-Square (χ^2)

For a cross-table, the *row marginal* is:

$$R_y = \sum_{k_X} n_{yx}$$

and the *column marginal* is:

$$C_x = \sum_{k_Y} n_{yx},$$

The expected number of observations in each cell
if X and Y are unrelated is:

$$E_{yx} = \frac{R_y \times C_x}{N}. \quad (1)$$

Chi-Squared (continued)

Under the “null” hypothesis of the independence of Y and X , we would expect:

- On average $n_{yx} = E_{yx}$.
- The difference between n_{yx} and E_{yx} will be small.

“Chi-Square”:

$$\chi^2 = \sum_{k_Y k_X} \frac{(n_{yx} - E_{yx})^2}{E_{yx}}.$$

Chi-Square, Explained

- Larger values of chi-square = greater deviation from “independence” of X and Y
- By itself, chi-square doesn't have any intrinsic meaning
- Used as a statistical test for whether or not two variables are associated

Chi-Square: Africa Example

```
> chisq.test(TABLE)
```

```
Pearson's Chi-squared test with Yates'  
continuity correction
```

```
data: TABLE
```

```
X-squared = 0.090515, df = 1, p-value = 0.7635
```

```
Warning message:
```

```
In chisq.test(TABLE) : Chi-squared approximation  
may be incorrect
```

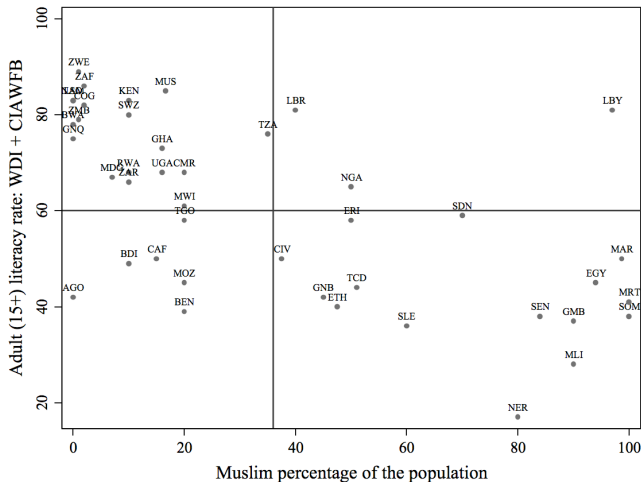
Statistical Association and Levels of Measurement

Some Measures of Association

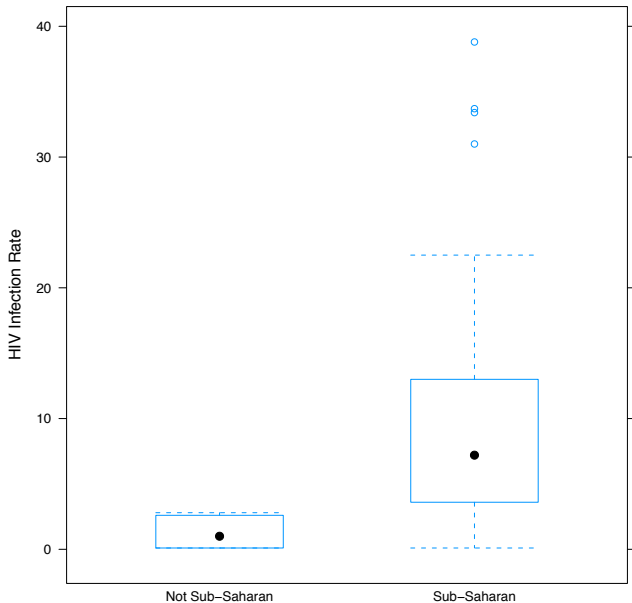
		X			
		Nominal	Binary	Ordinal	Interval/Ratio
Y	Nominal	χ^2	χ^2	χ^2	t -test (and η)
	Binary	χ^2	ϕ, Q	γ, τ_c	t -test
	Ordinal	χ^2	γ, τ_c	γ, τ_a, τ_b	Spearman's ρ
	Interval / Ratio	t -test (and η)	t -test	Spearman's ρ	r (+ regression)

Graphics are best. Period.

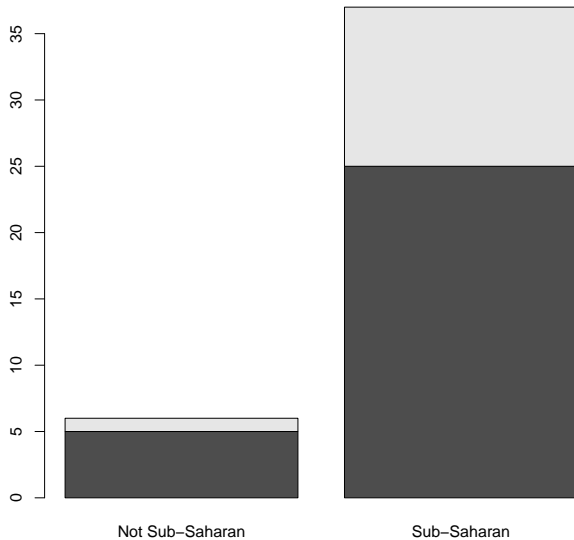
Two-Way Graphs: Scatterplots



Conditioned Boxplots



Africa Example: Barplot



Assessing association:

- In general, graphs $>$ tables
- Use graphs to convey size, shape, and strength of relationships
- Use statistics to test for “statistical significance”
- Beware (sometimes) hidden statistical assumptions (e.g., monotonicity, linearity)