02 -Review of Statistics and Demography

Ness Sandoval

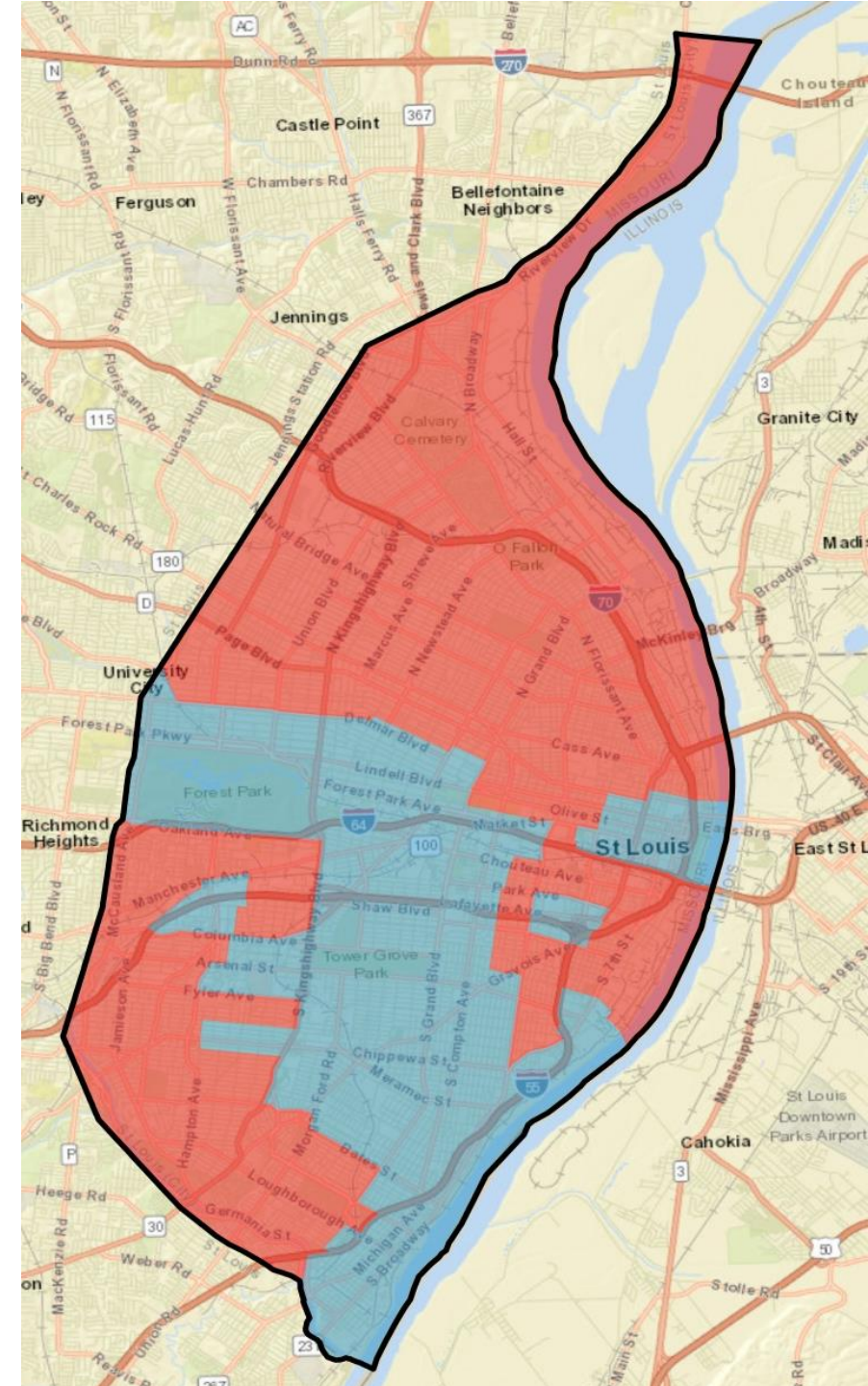Sociology

Saint Louis University

Outline

- Measurement

- Descriptive Statistics

- Type I and Type II

- Simple Linear Regression

- Advanced Topics in Regression

- Review of Demography

- Example

Contributes to Integration

Contributes to Segregation

Measurement

Measurement

	Nominal	Ordinal	Interval	Ratio
Examples	Categories	Ranks	Test scores and scales	Weight, number of responses
Properties	Identity	Identity Magnitude	Identity Magnitude Equal Interval	Identity Magnitude Equal Interval True Zero
Mathematical Operations	None	Rank order	Add; subtract	Add, subtract; multiply and divide
Type of data	Nominal	Ordered	Score	Score

Descriptive Statistics

Measures of Central Tendency

Numbers that describe what is average or typical of the distribution.

- The **Mode** is the category or score with the largest frequency in the distribution

- The **Median** is the score that divides the distribution into tow equal parts so that half the cases are above it and half below it.

- The **Mean** is the arithmetic average obtained by adding up all the scores and dividing by the total number of score

Measure of Variability

- The *variance* is a measure of variation for interval-ratio variables; it is the average of the squared deviation from the mean.

- The *standard deviation* is a measure of variation for interval-ratio variables; it is equal to the square root of the variance.

Shape of the Distribution

Symmetrical distribution
- The frequencies at the right and left tails of the distribution are identical; each half of the distribution is the mirror image of the other.

Skewed distribution
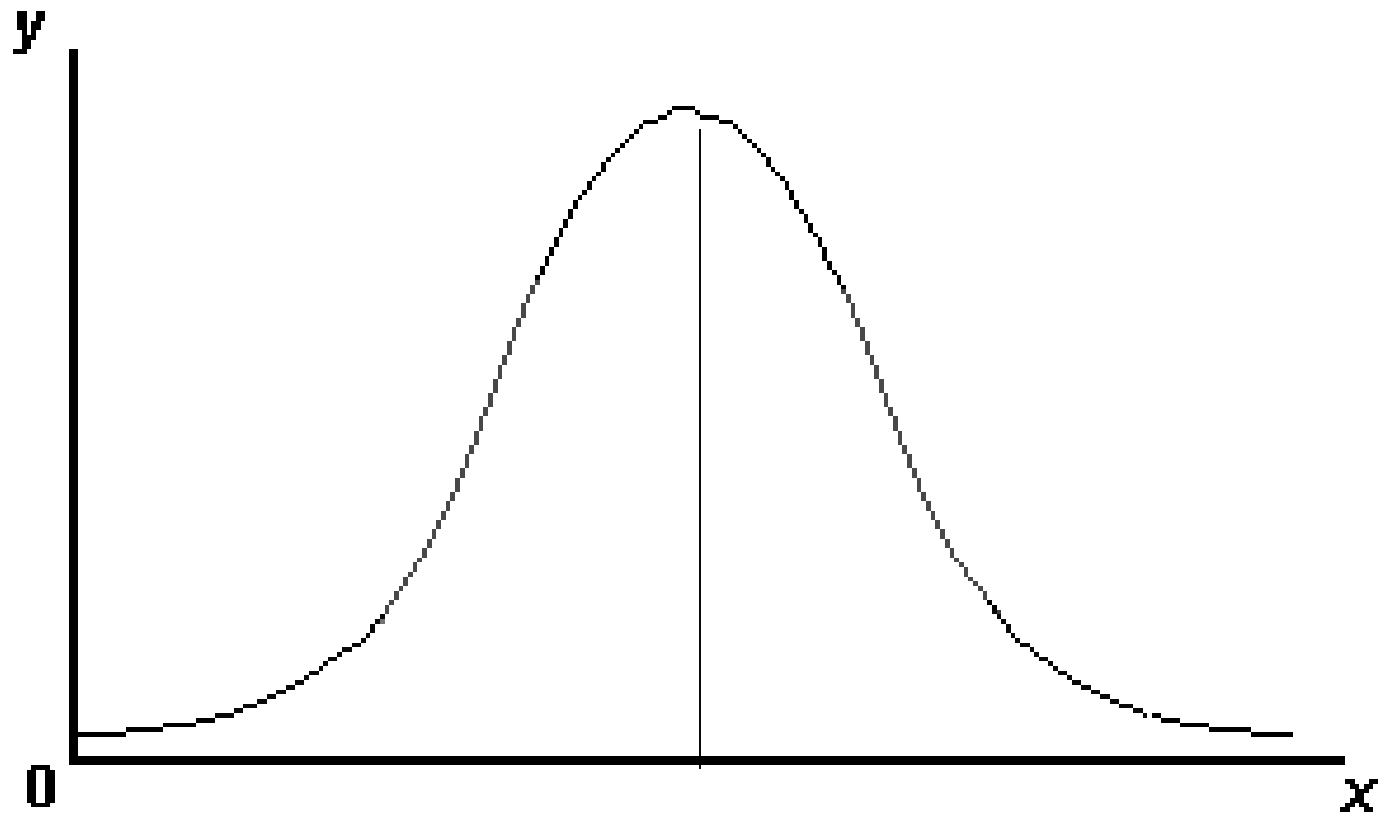- A distribution with a few extreme values on one side of the distribution.

Negatively skewed distribution
- A distribution with a few extremely low values.
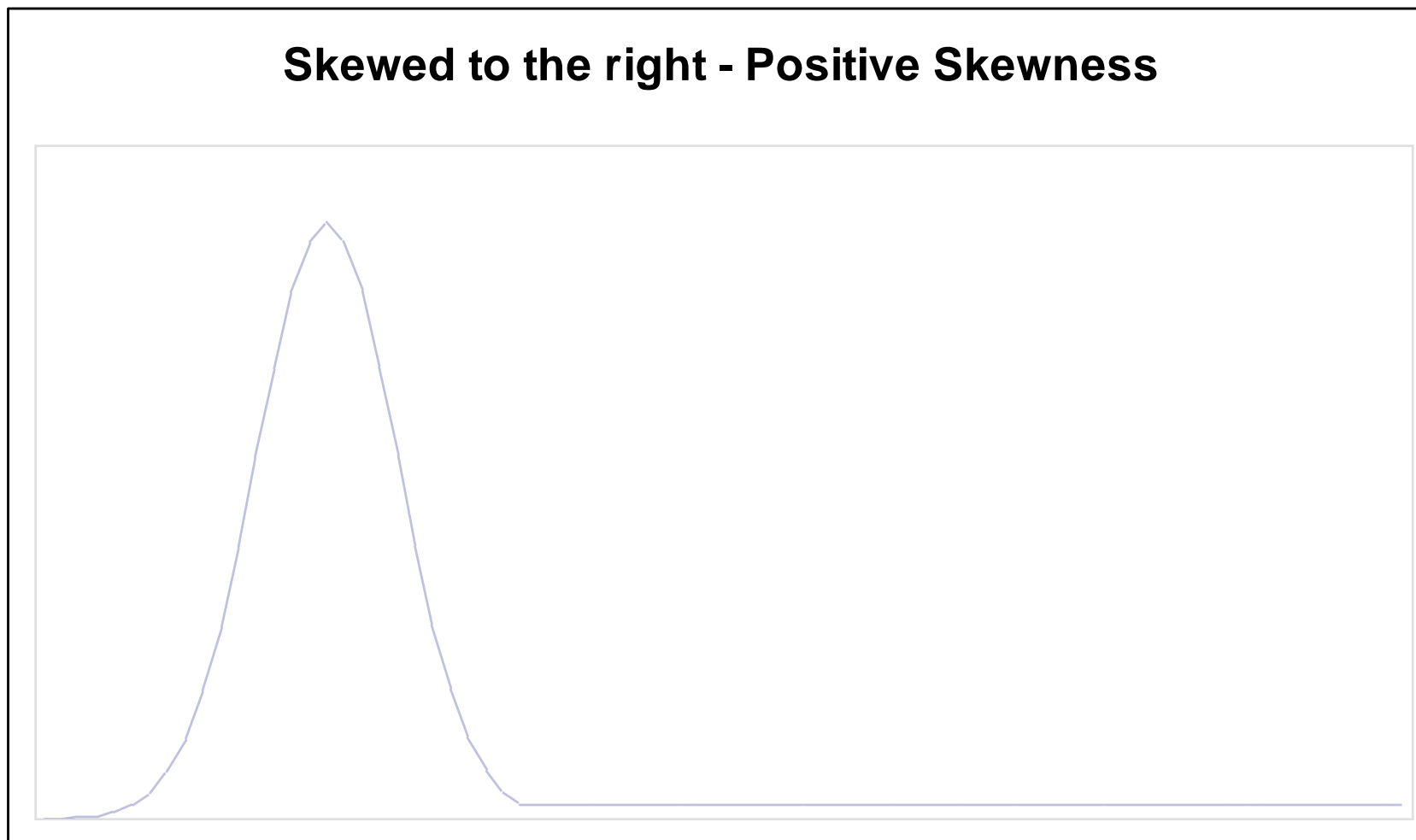
Positively skewed distribution
- A distribution with a few extremely high values.

The Normal Curve

Mean=Mode=Median

Skewness in the Distribution

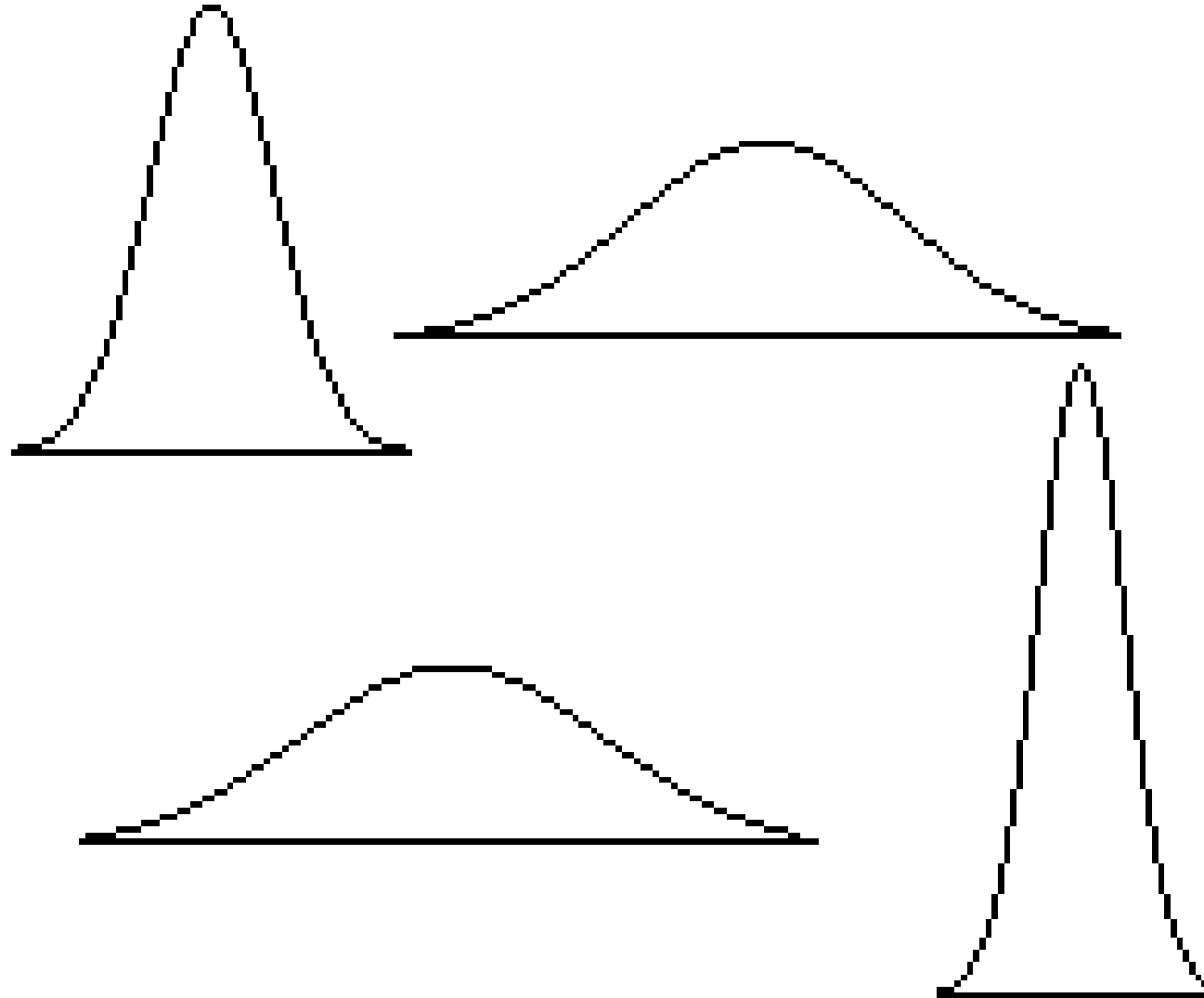**Skewed to the right - Positive Skewness**

Mean>Median>Mode

Skewness in the Distribution

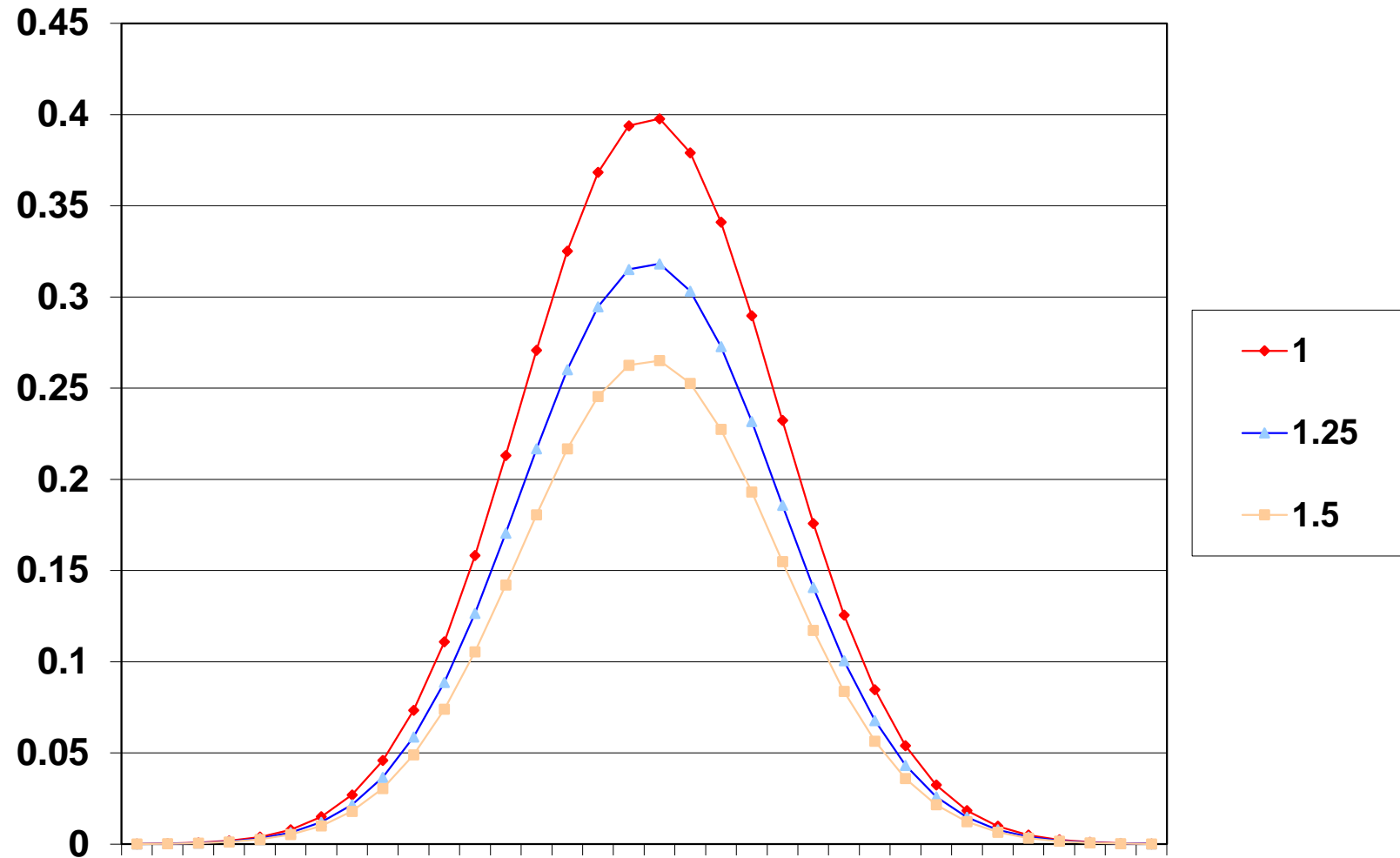**Skewed to the Left - Negative Skewness**
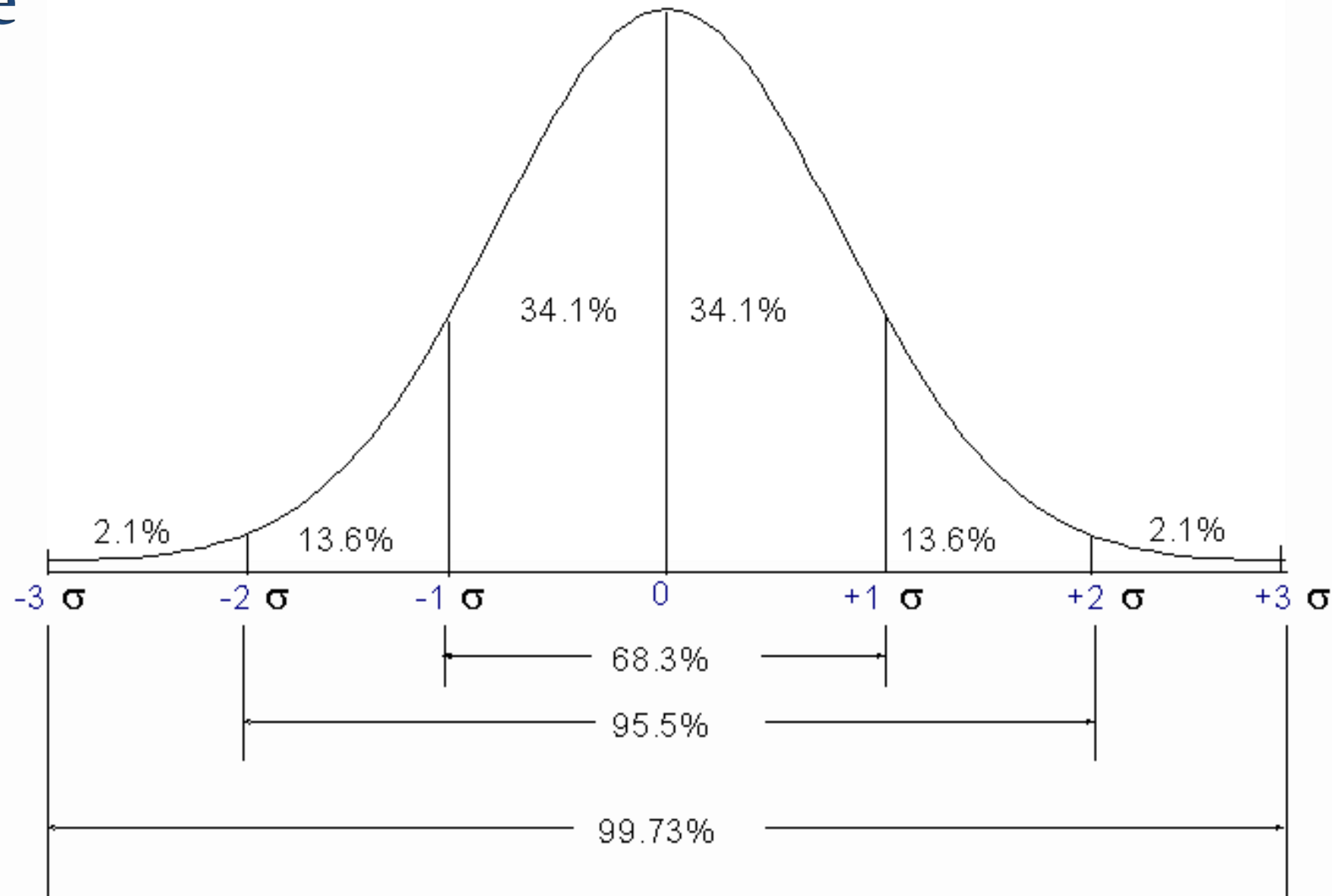

Mean<Median<Mode

Characteristics of a Normal Curve

Normal Distribution with Equal Means but Different Standard Deviations

Percentages Under the Normal Curve or "The 68-95-99 rule"

Type I and Type II Error

Errors in Hypothesis Testing

- *Type 1 Error* – The probability associated with rejecting a null hypothesis when it is true.

- *Type II Error* – The probability associated with failing to reject null hypothesis when it is false.

	H_o is True	H_o is False
Reject H_o	**Type I Error**	Correct Decision
Do not reject H_o	Correct Decision	**Type II Error**

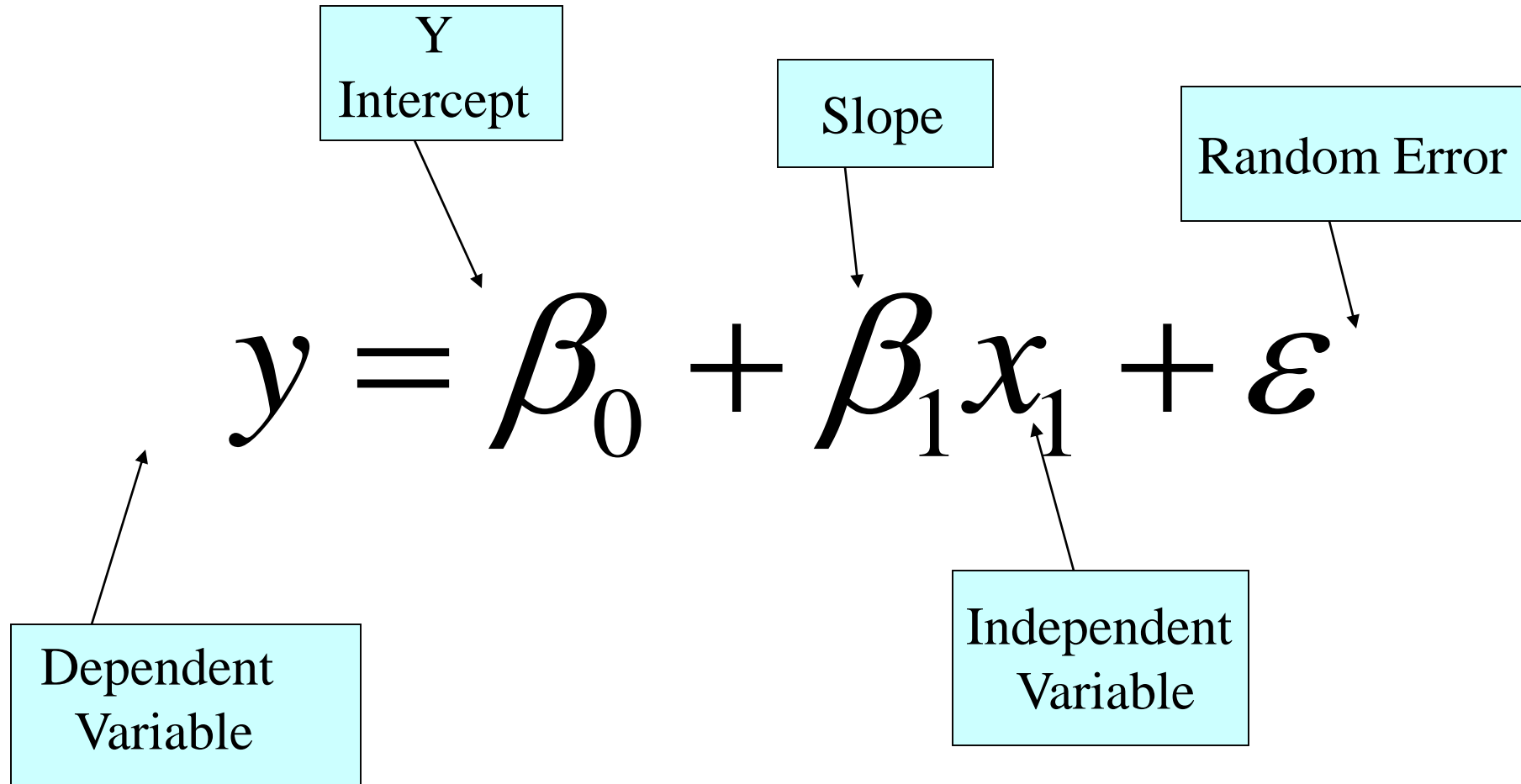# Simple Linear Regression

This looks better!

```
. corr  rincom98  sex age  white black other
(obs=946)

             | rincom98        sex        age      white      black      other
-------------+------------------------------------------------------------------
    rincom98 |   1.0000
         sex |  -0.2383     1.0000
         age |   0.2117    -0.0420     1.0000
       white |   0.1580    -0.0758     0.1337     1.0000
       black |  -0.1351     0.0766    -0.0412    -0.7562     1.0000
       other |  -0.0675     0.0175    -0.1502    -0.5538    -0.1261     1.0000
```

The Regression Model

Y
Intercept

Slope

Random Error

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Dependent
Variable

Independent
Variable

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

- ε= the deviation of the value Y from the mean value of the distribution given X.

  This error term may be conceived as representing:
  1. The effects on Y of variables not explicitly included in the equation

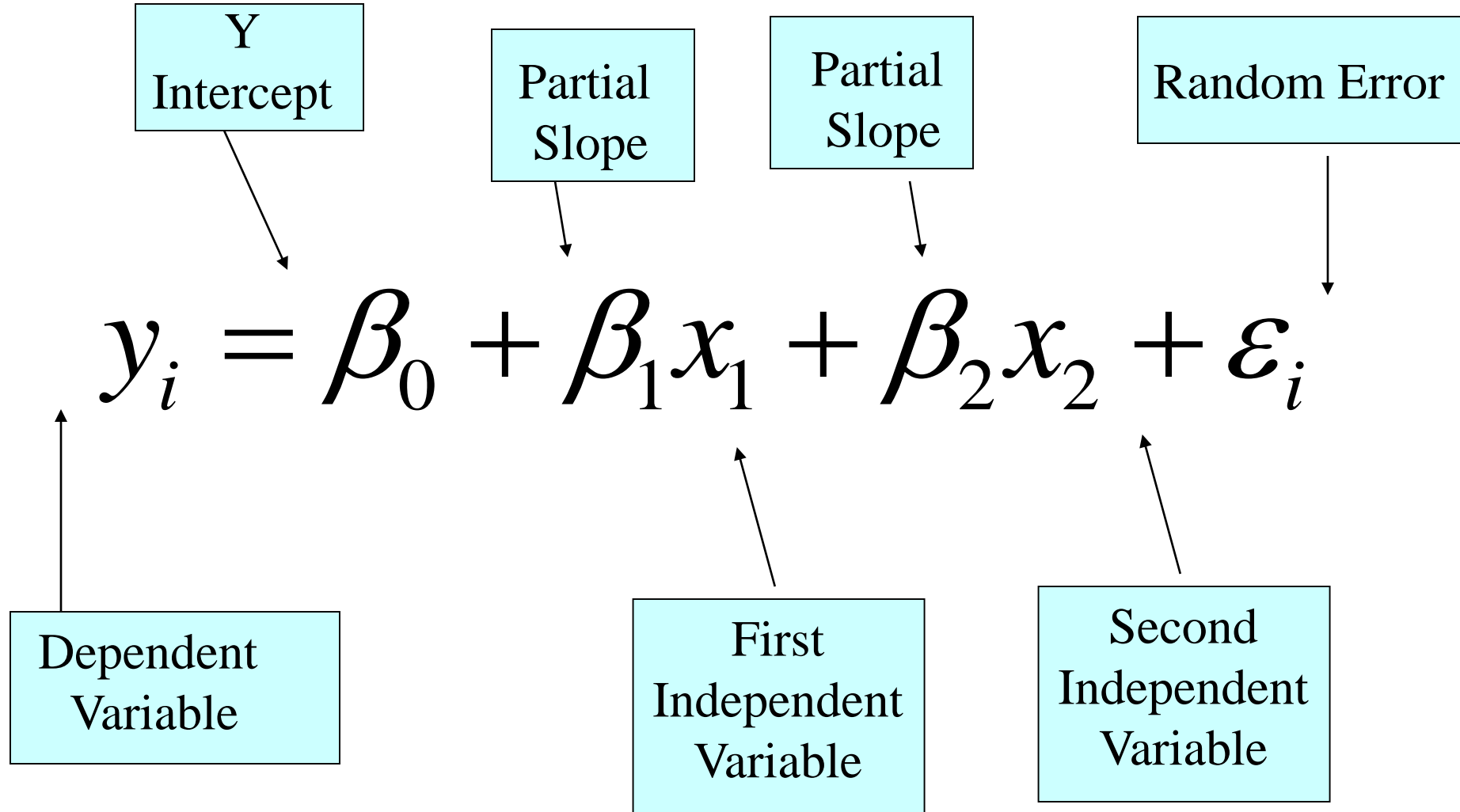 2. A residual random element in the dependent variable.

ε

- For historical reasons, the most popular philosophy about ε is that it is "random noise." Some researchers also call it "disturbance" or "unobserved."

  More specifically, that it is:
  1. uncorrelated with other variables

  2. has a mean value of zero.

Advanced Topics In Regression

The Regression Model- Two or More Independent Variables

Y Intercept

Partial Slope

Partial Slope

Random Error

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$$

Dependent Variable

First Independent Variable

Second Independent Variable

BLUE
Best Linear Unbiased Estimator - Part 1

R1. **Linear In Parameters**

R2. **Random Sampling**

R3. **All variables are measured without error**

R.4 **Zero Conditional Mean**

R5. **No Perfect Collinearity**

R6. **Homoskedasticity**

Violation	Impact						
	F	R-square	β	Std Error of Estimate	Std Error of β	T	Number of Violations
Measurement Error in Dependent variable	👍	👍	👍	👍	☹↑	☹↓	2
Measurement Error in Independent variable	☹	☹	☹	☹	☹	☹	6
Irrelevant Variable	👍	👍	👍	👍	☹↑	☹↓	2
Omitted variable	☹	☹	☹	☹	☹	☹	6
Incorrect functional form	☹	☹	☹	☹	☹	☹	6
Heteroskedasticity	☹	👍	👍	☹	☹	☹	4
Collinearity	👍	👍	👍	👍	☹↑	☹↓	2
Simultaneity Bias	☹	☹	☹	☹	☹	☹	6

Legend for Table

👍 = The statistic is still reliable and unbiased.

☹ = The statistic is biased, and thus cannot be relied upon.

↓ = Downward bias in estimation.

↑ = Upward bias in estimation.

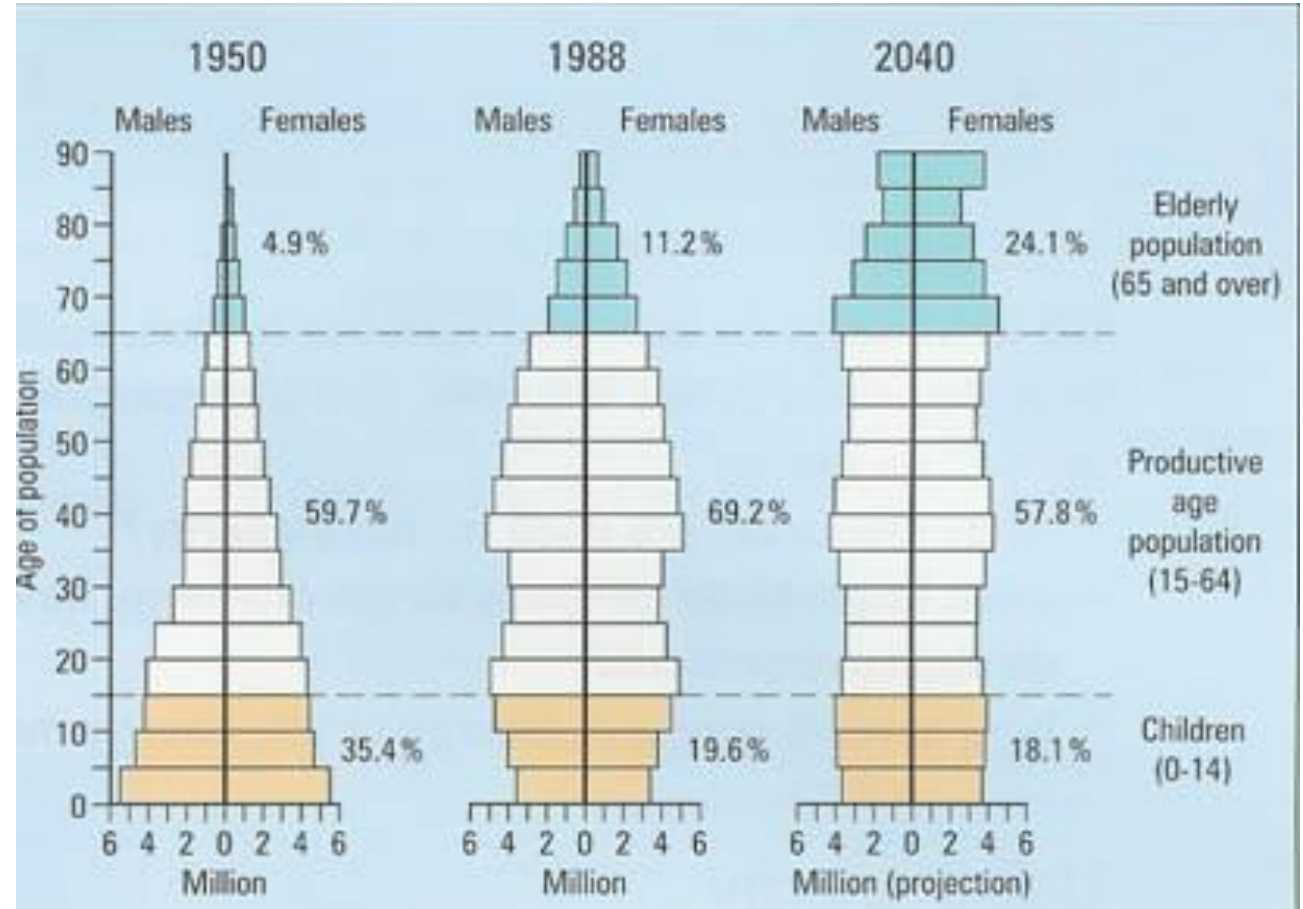# Review of Basic Concepts in Demography

What is Demography?

- Demography is the scientific study of human populations.

  - 1855 (Achille Guillard)

    - Demos – people

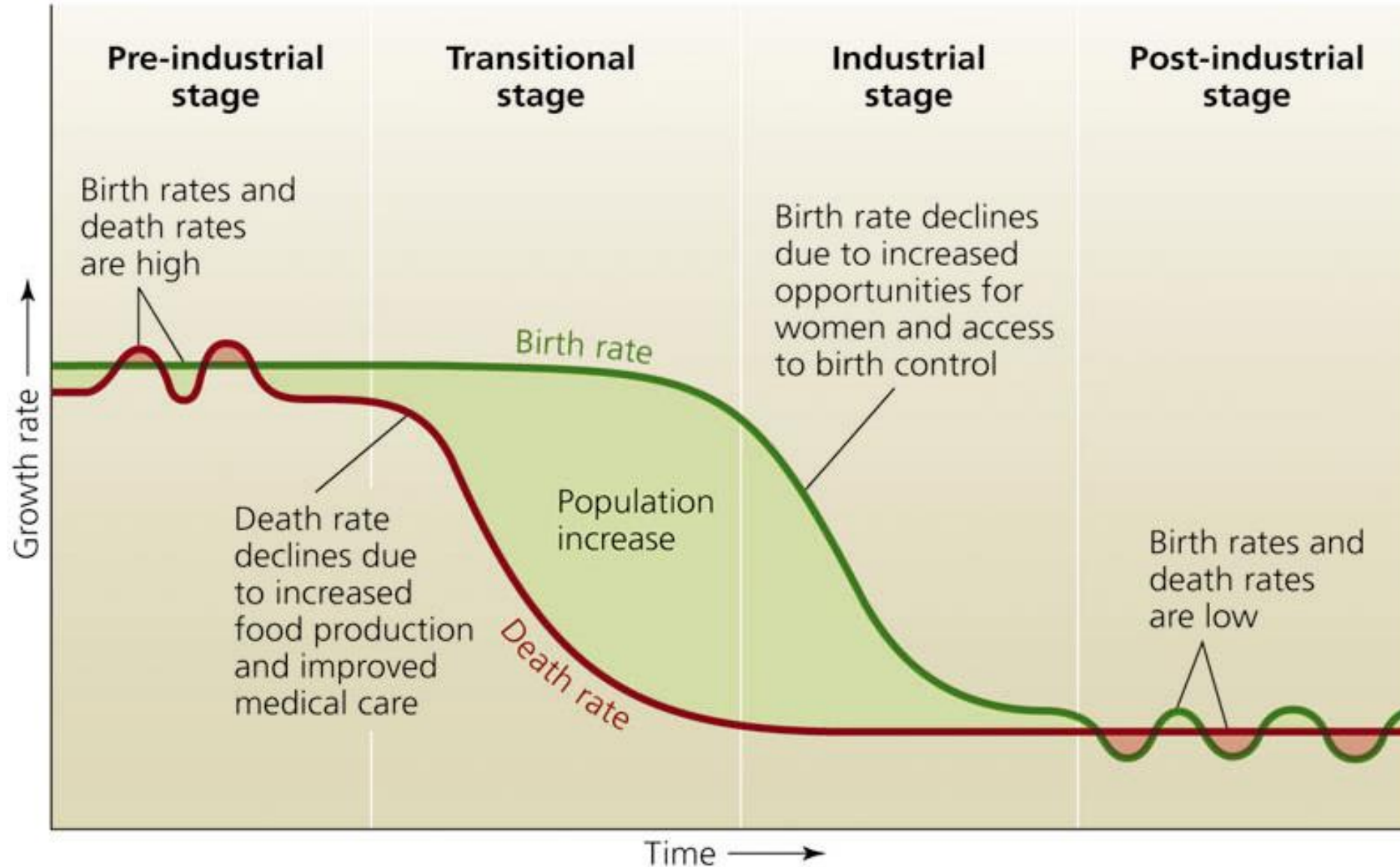   - Graphein – write about a particular subject

What is Demography?

Mathematical knowledge of **populations,** their general movement of the **population**, and their physical, civil, intellectual and moral state.
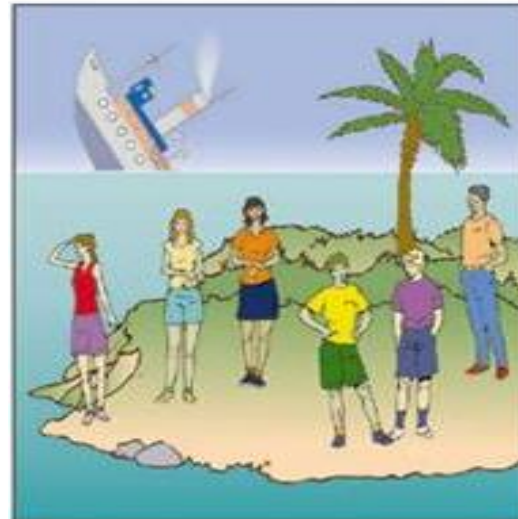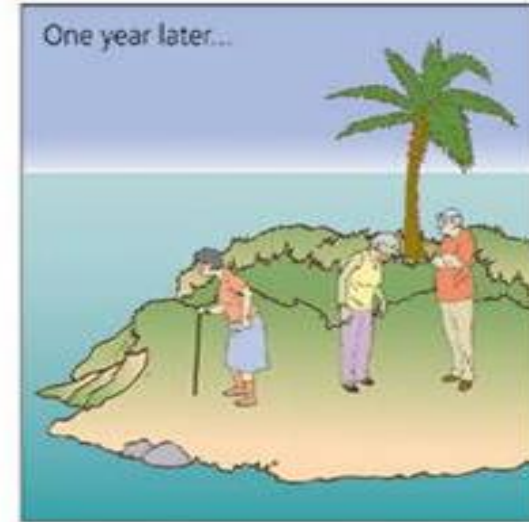
Demographic transition: Stages
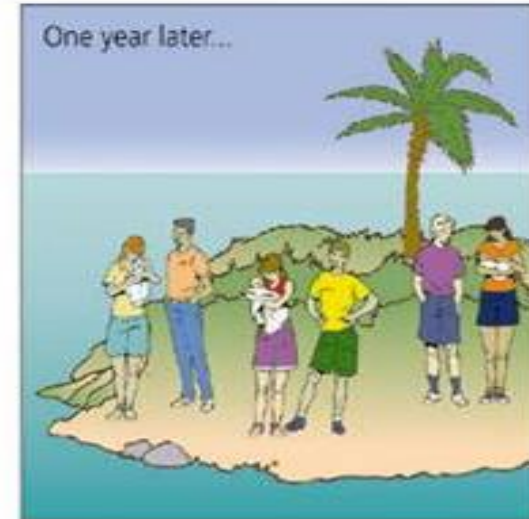
Age structure

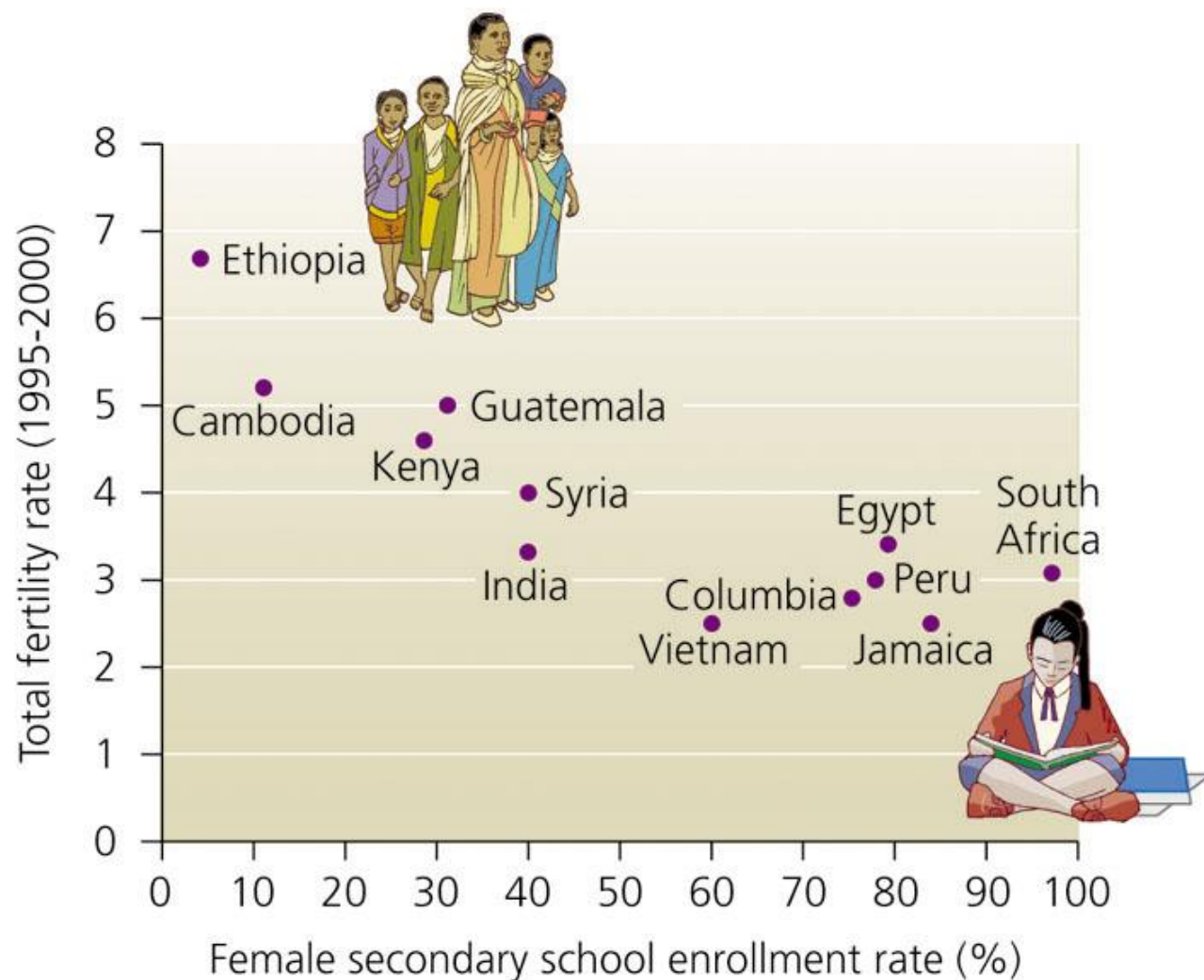Age structure can influence population growth rates.

(a) An older age structure

(b) A younger age structure

Female education and TFR

Female literacy and school enrollment are correlated with total fertility rate:

More-educated women have fewer children.

Ecological footprints

Residents of some countries consume more resources—and thus use more land—than residents of others.

Shown are **ecological footprints** of an average citizen from various nations.
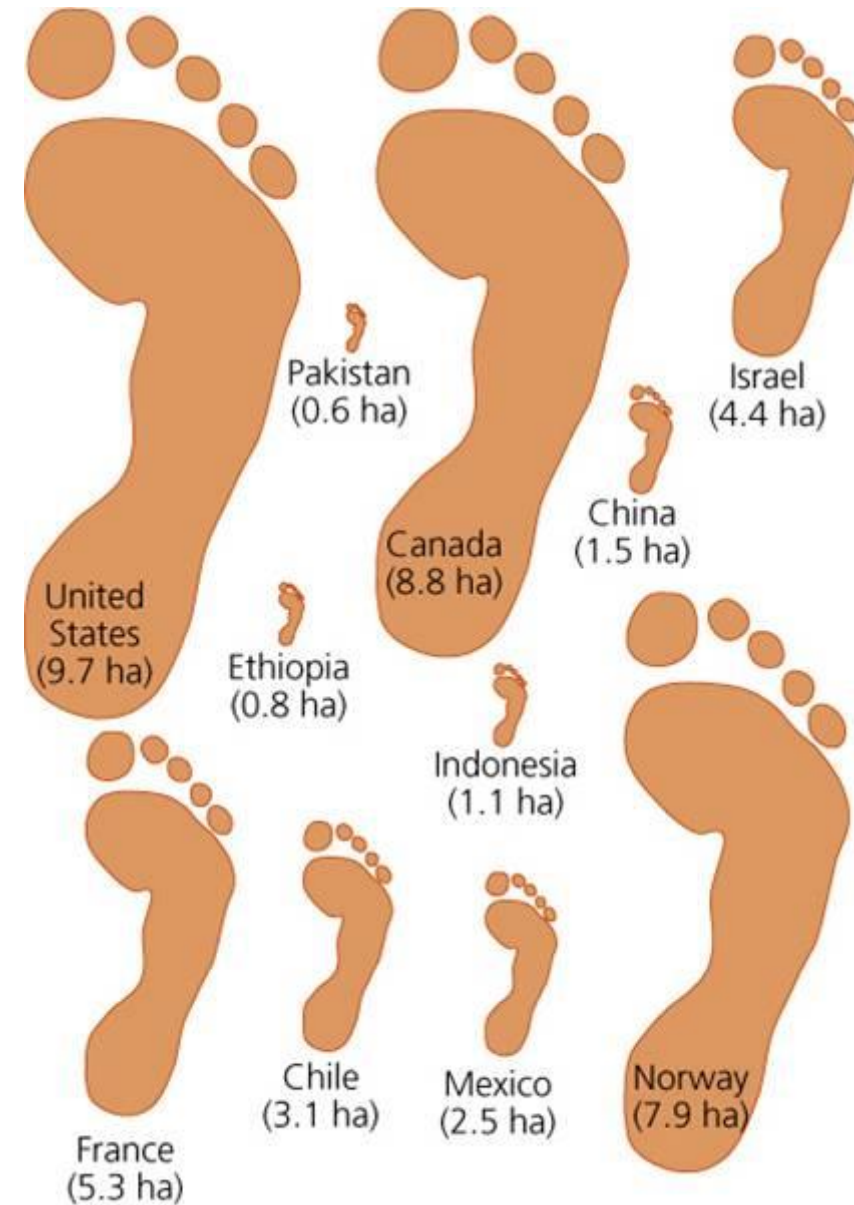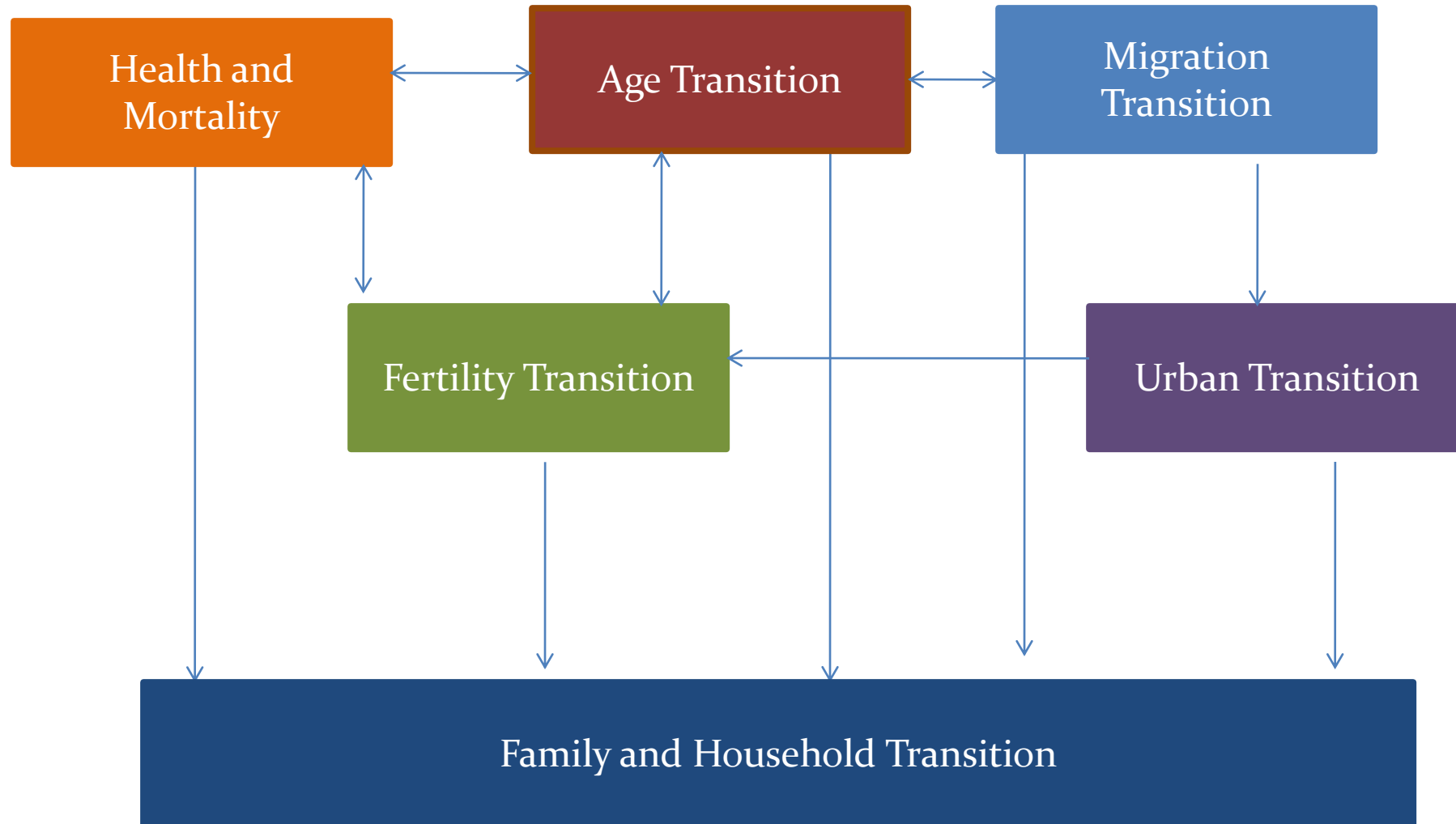

Pakistan (0.6 ha)

Israel (4.4 ha)

Canada (8.8 ha)

China (1.5 ha)

United States (9.7 ha)

Ethiopia (0.8 ha)

Indonesia (1.1 ha)

Chile (3.1 ha)

Mexico (2.5 ha)

Norway (7.9 ha)

France (5.3 ha)

Figure 7.23

The "IPAT" model

Shows how **P**opulation, **A**ffluence, and **T**echnology interact to create **I**mpact on our environment.

$$I = P \times A \times T$$

Decomposition of the Demographic Transition Model

Example

Percent Black

LISA Map
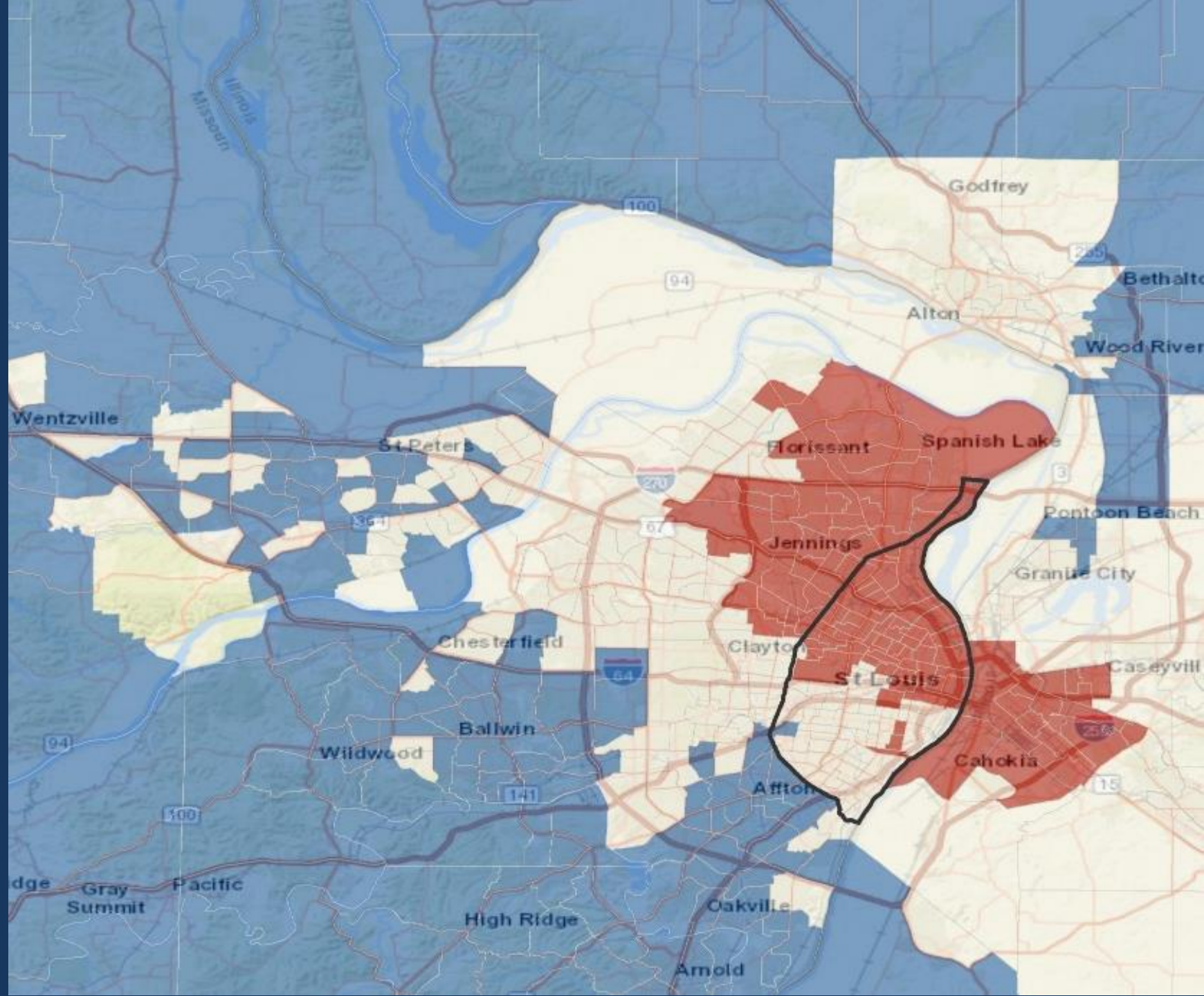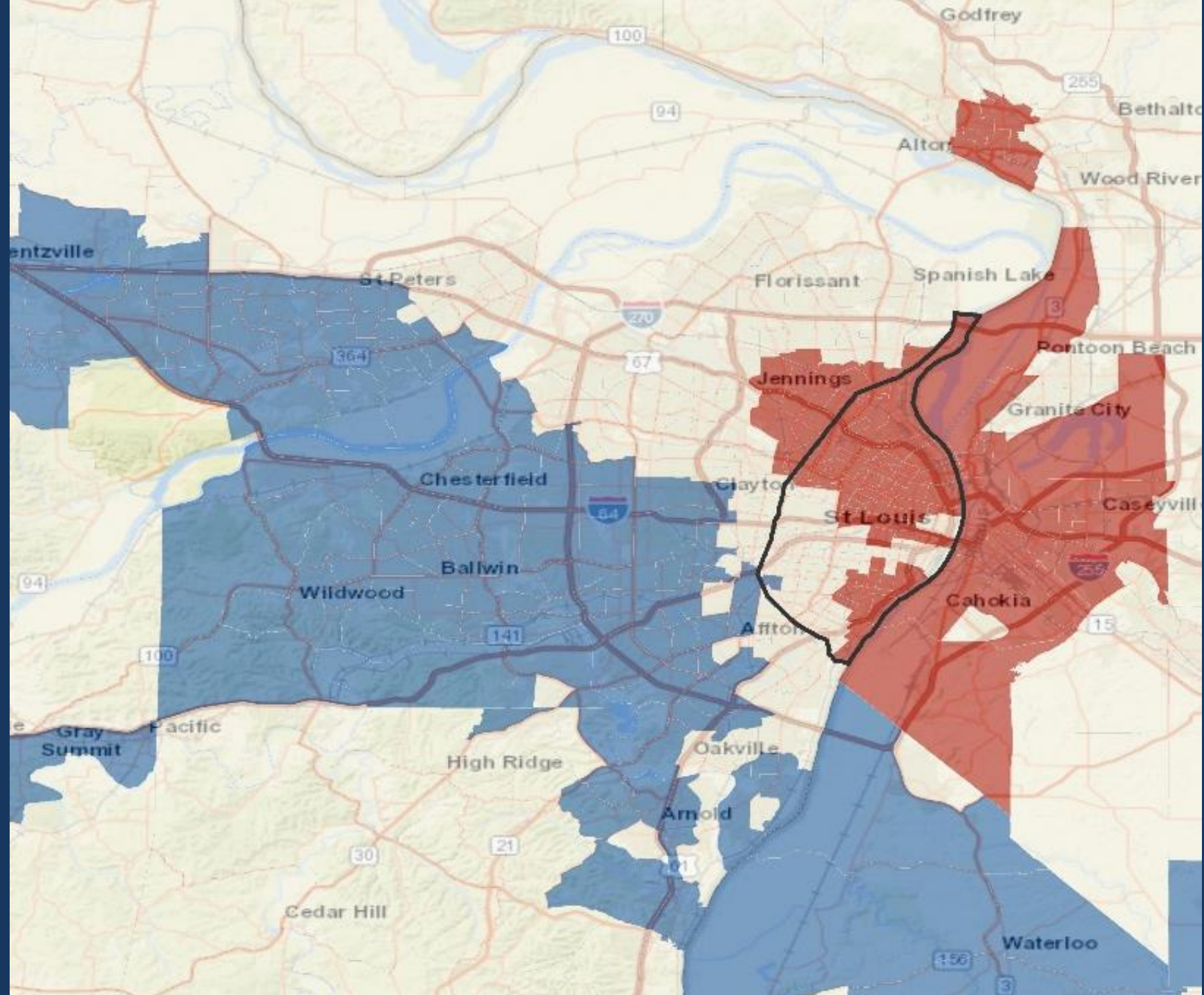
⬜	Not Significant
🟥	High-High Cluster
🟦	Low-Low Cluster

Poverty Rate

LISA Map

- ⬜ Not Significant
- 🟥 High-High Cluster
- 🟦 Low-Low Cluster

Poverty Rate and Percent Black

BiLISA Map

High-High Cluster

Low-Low Cluster