Min-Seok Pang
Management Information Systems
Fox School of Business, Temple University
minspang@temple.edu
Sep. 8th, 2015

MIS 0855 Fall 2016 – Data Science
Week 2 – In Data We (Mis)Trust
Do you trust your data?

- You have a decision to make.
- You have theory in mind and hypotheses to test.
- You have data at your hand!
- Do you believe in your data?

Can we trust the FBI’s national crime stats?

Uniform Crime Reports

About Crime in the U.S. (CIUS) By offense, by region, by state, by local agency

Offenses Known to Law Enforcement
- Violent Crime
- Property Crime
- Clearances
  (offenses "closed" by arrest or exceptional means)
- Go to Offense Tables

Persons Arrested
- Expanded Offense Data
  - Homicides
  - Trends
  - Rates
  - Weapons
- Includes data about the age, gender, and race of arrestees for 28 separate offenses including murder
  - Go to Arrest Tables

Police Employee Data
- Includes data about sworn officers and civilian employees
  - Go to Police Employee Tables

THE BASICS

What is the UCR Program?

The UCR Program is a voluntary city, university and college, county, state, tribal, and federal law enforcement program that provides a nationwide view of crime based on the submission of statistics by law enforcement agencies throughout the country.

OFFENSE INFORMATION

Which specific crimes are reported to the UCR Program, and why were these crimes identified for reporting?

The UCR Program collects offense information for murder and nonnegligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, and arson. These are Part I offenses and are serious crimes by nature and/or volume. Not all crimes, such as embezzlement, are readily brought to the attention of the police. Also, some serious crimes, such as kidnapping, occur infrequently. Therefore, the UCR Program limits the reporting of offenses known to the eight selected crime classifications because they are the crimes most likely to be reported and most likely to occur with sufficient frequency to provide an adequate basis for comparison.

http://www.fbi.gov/about-us/cjis/ucr/frequently-asked-questions/ucr_faqs
Problems with UCR Data

- The UCR reports only crimes known to the police.
- The UCR reports on only the most serious crime incidents.
- The UCR does not collect all relevant data.
  - “Forcible rape of a male victim is recorded as an aggravated assault.”
- The UCR reveals more about police behavior than it does about criminality.
  - Reporting by local police to the FBI is on a voluntary basis.

Unemployment Rates – Can We Trust This?

Source: Bureau of Labor Statistics

How to Calculate Unemployment Rates

- Unemployment rates can be lower when more people stop looking for work.
- When more people work for part-time jobs.

Unemployment Rate = \frac{\text{Unemployed}}{\text{Labor Force}} \times 100

People looking for work
People working
People looking for work
Washington D.C. Potholes Map
(http://ddot.dc.gov/service/pothole-repair/)
Reported via DC 311 Mobile App


Source: U.S. Census Bureau, 2006–2010 American Community Survey. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see <www.census.gov/acs/www>.

Do we have to believe Yelp stars and reviews?


"Even when I moved to a new apartment farther away I will still take the extra walk to go there instead of the other bars near temple," in 5 reviews

"EDIT: Whalen just made me a cocktail based on the following directions "I like tequila, I don't like sweet, and I love spicy" and then he made me a cocktail sent from the tequila gods. :)" in 2 reviews

"Plenty of TVs on the wall and the staff are cool about switching it to whatever sporting event you may be interested in." in 3 reviews

Has TV: Yes