ALL ABOUT STATISTICS
TECH Workshop

TUESDAY • MAR 7 • 12:30 • L315
Register online: guides.law.fsu.edu/TechTuesdays

Free Lunch!

FLORIDA STATE UNIVERSITY | COLLEGE OF LAW RESEARCH CENTER
Epidemiology
Epidemiology is the branch of medicine that deals with the incidence, distribution, and possible control of diseases and other factors relating to health.
Epidemiology

- It’s possible to screen for diseases.
- Let’s say we have a test for a disease that is 99% accurate.
We test one person randomly.
They test positive.
What are the odds they have the disease?
Rarity of the disease

- It’s a rare disease, only 1 in 10,000 people have the disease.
Rarity

You test one million people. At that rarity 100 people would have the disease and since the test is 99% accurate, it would correctly identify 99 people as having the disease and it would miss one.
Results out of a Million

![Bar chart showing results out of a million with a large positive bar and a small negative bar.](chart.png)
Rarity

- We tested one million people for the disease, and we know already that it will correctly identify 99 of the 100 people that have the disease.
- All but 100 people will be negative for disease.
Rarity

You test one million people for the disease, we know already that it will correctly identify 99 of the 100 people that have the disease, the overwhelming majority of the 1,000,000 people will
Incorrectly Diagnosed

- But since the test is 99% accurate, it will falsely identify about 10,000 people as having the disease who do not.
Sally Clark, mother wrongly convicted of killing her sons, found dead at home

- Family says she never recovered from court case
- Cause of death to be determined by coroner

Sally Clark

- Clark’s first son died a few weeks after birth
- Clark’s second son also died shortly after birth
- Both cases appeared to be from Sudden Infant Death Syndrome (SIDS)
A pediatrician testified that two children dying from SIDS from an educated and affluent family (which the Clark family was) was very unlikely.
He concluded, SIDS occurred in non-smoking healthy household at a rate of about 1/8500
Pediatrician

That two SIDS death would occur at a rate of $\frac{1}{8500} \times \frac{1}{8500}$ or about 1 in 73,000,000 births.
Flipping a Coin

- The odds of getting a heads or tails when you flip a coin is 50% or $\frac{1}{2}$.
- The odds of getting two heads in a row are $\frac{1}{2} \times \frac{1}{2}$ or 25%. 
That two SIDS deaths would occur at a rate of 1/8500 X 1/8500 or about 1 in 73,000,000 births.
The Jury

- It was presented to the jury that the chances that she was innocent was 1 in 73,000,000
Convicted

- Sally Clark was convicted
Are there any statistical problems in this analysis?
First Problem

- One problem is the rareness of parents who kill their children.
- This was not presented to the jury.
This is essentially the same problem we saw with the epidemiological problem of testing for a disease that is 99% accurate.
Rarity

- Two SIDS death in one family is very rare.
- Parents who kill their children is a very rare occurrence.
Odds of Guilt or Innocence

- The notion that the odds that she was innocent was 1 in 73,000,000 is just statistically wrong.
- But the presenting of that notion was not challenged during the trial.
Independence

- When I flip a coin the odds are 50-50 each time I flip the coin.
- Each flip is independent of the previous flip.
- 10 heads in a row is rare because the odds are 50-50 each time I flip the coin.
Independence
Independence

- No evidence of independence was presented at the trial.
- In fact there are both **genetic** and **environmental** factors that increase the risk of SIDS.
Environmental Factors

- In this case it was revealed on appeal that one of the children had a staph infection at the time of death which was probably contributory to the death.

- (Somehow at the original trial the prosecution was aware of this but the defense was not.)
Genetics

- Plus there are genetic factors.

SIDS: Genetics May Play a Role

Finding Helps Explain Why More Black Infants Die of SIDS

By Jeanie Lerche Davis

Aug. 19, 2004 -- The tragic mystery of sudden infant death syndrome (SIDS) is a bit closer to being solved. New research shows that genetics may play a role.
Genetics

- SIDS: Genetics May Play a Role

- Targeting Sudden Infant Death Syndrome

  - Known genetic disorders of metabolism likely account for up to 5 percent of SIDS cases. Other genetic diseases, such as prolonged QT syndrome, may contribute to SIDS in a small percentage of cases.

  - Increasing knowledge of SIDS-associated abnormalities, such as developmental brainstem anomalies or association with upper respiratory infections, is suggesting possible genetic risk factors (i.e., “susceptibility genes”) that may predispose an infant to SIDS.

  - Genetic factors or their interactions with environmental factors may account in part for the variations in SIDS incidence observed across subgroups of the population.
After three years in prison her second appeal was successful and she was released.
Not a Happy Ending

- Unfortunately this does not have a happy ending.
Misreading Statistics

In this case had tragic consequences.
Very Basic Examples of Misleading Statistics
Massive Increase in Home Price

Average House Price in €:

<table>
<thead>
<tr>
<th>Year</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>80000</td>
</tr>
<tr>
<td>1999</td>
<td>82500</td>
</tr>
</tbody>
</table>
Massive Increase in Home Price

Average House Price in €

1998  1999
Comparison

http://www.mathcaptain.com/algebra/misleading-graphs.html
The Navy and Safety

Many years ago

- The Navy used a statistic that on average nine sailors out of thousand died a year while in the Navy.
- While the civilian population of New York city, the number was 16 out of thousand.
- The conclusion, it was safer to be in the Navy.
Once a police office told me that he thought Mercedes drivers were the best drivers, because he almost never saw a Mercedes in a car accident.
Wine

More subtle: It was widely reported some years ago that drinking wine may increase life expectancy.

This was based on research that compared moderate wine drinkers with non-drinkers.

On the surface it seems reasonable to accept that but...
Wine

- It wasn’t established among the non-drinkers why they didn’t drink.
  - Kidney disorders
  - Liver disorders
  - Medications
Illuminating Innumeracy

- Professor Laurence Tribe has suggested that jurors might overvalue numerical data.
- But others have argued that statistical information that’s difficult to process, tends to be ignored.

Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*
Finding Data
Data Resources and Statistics for Legal Research

Introduction

We are surrounded by data but what are data? Data are essentially numbers that haven't been analyzed. There are probably thousands of datasets that are online. This guide includes links to online sources for gathering data focusing in particular on government datasets. To study the data, you need to develop a hypothesis, find the data you need, and then analyze the data. This guide also includes links to books and articles on empirical research, content to increase current awareness, it covers software resources for statistical analysis and where you can find help on campus for empirical research.
US Government Statistics

U.S. Government Statistics

- Bureau of Labor Statistics
- Bureau of Economic Analysis
- CDC - National Center for Health Statistics
- Census Bureau
- Equal Employment Opportunity Commission
- Environmental Protection Agency
- Fed. Reserve - Economic Research & Data
- Homeland Security - Immigration Data & Statistics
- IRS - Tax Statistics
- Occupation Safety and Health Administration
- Office of Policy Development and Research (HUD)
- Social Security
- DataRefuge
- Internet Archive

- Economic Research & Data
- Data Releases
- Data Download Program
- Research Resources Survey of Consumer Finances

U.S. recessions

Log Scale | Style | Linear | Fullscreen | Save Image | Save PDF
Data Refuge and WayBackMachine

**DataRefuge**

DataRefuge helps to build refuge for federal data and supports climate and environmental research and advocacy. We are committed to fact-based arguments. DataRefuge preserves the facts we need at a time of ongoing climate change.

**Wayback Machine**

Founded in 1996, the mission of the Internet Archive is to provide Universal Access to All Knowledge. It collects published works and makes them available in digital formats. The Internet Archive is building a public library that can serve anyone in the world with access to the Internet.
Foreign and International Datasets

- Armed Conflict Location & Event Data Project
- Correlates of War Project
- ICOW - Project description

Map of Political Violence in Africa and Asia 2015
Foreign and International Datasets

- UNdata
- European Data Portal
- Eurostat
- OECD Data
- International Debt Statistics
IBM Punch Cards

Until the mid-1970s, most computer access was via punched cards. Programs and data were punched by hand on a key punch machine and read into a card reader. Large computing sites such as Columbia University purchased cards by the truckload and furnished them free of charge to users. During the IBM 360 era (1969-80) Columbia’s cards were embossed with the legend "CUCC 360" (Columbia University Computer Center IBM 360) and the Columbia shield (In Lumine Tuo Videbimnus Lumen).

This is a punched card allowing color.

Links:
- Punch card
- Alles over ponskaarten: "Wat is een ponskaart en wat doe je er mee?", Philips, 1967 (Dutch)
- Everything about punch cards: "What is a punchcard... What does it do for me?" (English)
- Remington Rand Punch Card System
Card Punch Machine
IBM Punch Cards

Until the mid-1970s, most computer access was via punched cards. Programs and data were punched by hand on a key punch machine and read into a card reader. Large computing sites such as Columbia University purchased cards by the truckload and furnished them free of charge to users.

During the IBM 360 era (1969-80) Columbia's cards were embossed with the legend "CUCC 360" (Columbia University Computer Center IBM 360) and the Columbia shield (In Lumine Tuæ Videbimus Lumen).

Links:
- Punch
- Alles over ponskaarten: "Wat is een ponskaart en wat doet je er mee?", Philips, 1967 (Dutch)
- Everything about punch cards: "What is a punchcard... What does it do for me?" (English)
- Remington Rand Punch Card System
Statistics Software

- **SPSS**
  SPSS supports advanced statistical analysis and includes predictive analytics.
- **SAS**
  SAS is statistical programming software which supports analysis of variance and linear regression, statistical visualization, modeling, forecasting and much more.
- **STATA**
  STATA is statistical package designed for researchers of all disciplines.
- **R-Project for Statistical Computing**
  R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.
Software

- **SPSS**
  - SPSS supports advanced statistical analysis and includes predictive analytics.

- **SAS**
  - SAS is statistical programming software which supports analysis of variance and linear regression, statistical visualization, modeling, forecasting and much more.

- **STATA**
  - STATA is statistical package designed for researchers of all disciplines.

- **R-Project for Statistical Computing**
  - R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.
Academic Discount

IBM SPSS Statistics Base Grad Pack 24.0 Academic (Windows Download - 12 Month License)
Product ID: 1610290 | Mfg Part #: D0EWRLL_24
Who's Eligible (Details)
Student

SAVE $16,035.01
Retail: $16,300.00
$264.99

ADD TO CART

IBM SPSS Statistics Premium Faculty Pack 24.0 Academic (Mac Download - 12 Month License)
Product ID: 1610291 | Mfg Part #: D0EWTLL Mac_24
Who's Eligible (Details)
Faculty, Staff

Electronic Software Delivery
Data Visualizations

- FEMA - Data Visualizations
- U.S. Census Data Visualization Gallery
- NYTimes - Mapping America: Every City, Every Block
- The Data Visualization Catalogue
- The Use of Data Visualization in Government

HeinOnline

- Through a Glass Darkly: Using Brain Science and Visual Rhetoric to Gain a Professional Perspective on Visual Advocacy
- Visual Rhetoric: Topics of Invention and Arrangement and Tropes of Style

From SSRN

- Showing People behind Data: Does Anthropomorphizing Work?

Data Visualizations in Government

The report presents case studies on how visualization techniques are being used by two local governments, one state government, and three federal government agencies. Each case study discusses audience for visualization. Understanding audience is important when visualizing data. This report highlights how government organizations provide useful visualizations to different audiences, including the media, political oversight organizations, constituents, and internal program teams. To assist in efforts communicating to these audiences, the report details attributes of effective visualizations: relevance, meaning, beauty, ease of visualizing, truthfulness, accuracy, and consistency among them.

Tableau
Tableau Gallery – Oil Rigs
The gallery displays a visual representation of music hits from different decades. Users can select an artist and song from the drop-down menus. The chart shows the popularity of genres across the decades, with larger circles indicating more popular genres.

For example, the 1980s and 1990s are depicted with larger circles, suggesting these decades had popular music genres. The chart is not meant to compare genres across decades directly but to show the dominating genre in each.

Options for filtering genres and decade audio features are available, allowing users to customize their viewing experience.
I remember when I was a sophomore in high school and my friend told me that he just upgraded his computer to 32 mb of ram and I was totally blown away at how much ram that was. Now, 16 gb flash drives are sold in the “impulse buy” section in between batteries and candy bars.
Abstract

The legal brief of tomorrow will have color, graphics, photographs, imbedded videos, depictions, and diagrammatical elements; it will, in short, be highly visual. The tomorrow I am talking about is the very near future. The confidence with which I bring this prediction is bolstered by the research and scholarship of a wide range of disciplines — law and popular culture, cognitive studies and brain science, data visualization studies, and modern argument theory in rhetoric — and by the development of technology in the production of legal documents, and technology in the reading and reception of legal documents, that converge in one locus: that legal briefs will, and should, become a visual presentation of legal rhetoric and narrativity, and thereby fulfill the professional responsibility of advocates to use the best practices to represent the interests of their clients in law practice.
Workshop Tomorrow

MAR 8  LEARN HOW TO USE FALR & DOAH: THAT'S AN ORDER

_Instructor MARY McCORMICK_

Finding administrative rulings can sometimes feel like you are playing a game of hid-and-seek. In just 30 minutes, lose the frustration and learn how to find Florida administrative rulings and orders.
Data Visualizations