Data and Science

- Why do you think this course is called “Data Science”?
- What is science?
  - A collection of proven knowledge that
  - explains what is happening and
  - predicts what will happen.

http://escience.washington.edu/blog/uw-berkeley-nyu-collaborate-378m-data-science-initiative
Data, Information, and Knowledge

- For knowledge to be useful for informed decision making, it needs to be based on facts substantiated by data and information.
- So, how to build knowledge from data and information in a scientific manner?
- by devising formal hypotheses and testing them with data
A Hypothesis Should Be (1/2)

- **Testable** – be able to be measured by data
- **Falsifiable** – be able to be proven false by data
- **Grounded in a theory** – be able to be explained by sensible rationale(s)
A Hypothesis Should Be (2/2)

<table>
<thead>
<tr>
<th>Good</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Testable</strong></td>
<td>iPhone users download more apps each month than Android users.</td>
</tr>
<tr>
<td><strong>Falsifiable</strong></td>
<td>There are no vampires living in Louisiana.</td>
</tr>
<tr>
<td><strong>Grounded in a theory</strong></td>
<td>Students who attend class more often get better grades.</td>
</tr>
</tbody>
</table>
Any Theory or Rationale Behind This? (1/3)
Any Theory or Rationale Behind This? (2/3)

Fatherhood makes men FAT: New dads gain an average of 4lbs in the years after their first child is born

- Men who became fathers saw an average 2.6 per cent rise in their BMI
- For a 6ft man, is equivalent of putting on 4.4lb (2kg) from birth of first child
- However men who had no children actually lost around 1.4lb in weight

https://www.pinterest.com/pin/218072806933122983/
Any Theory or Rationale Behind This? (3/3)

Study Links Police Bodycams to Increase in Shooting Deaths

Create your own hypotheses (1/2)

● “Neighborhood A in Philadelphia is more dangerous to live in than Neighborhood B.”

● How would you make this more testable?

● What is the theory/rationale behind your hypothesis(e)s?

http://www.daniellemarks.com/Philly%20Neighborhoods.html
Create your own hypotheses (2/2)

● “Temple University is greater than Drexel University.”
● How would you make this more testable?
● What is the theory/rationale behind your hypothesis(es)?
In-Class Exercise (Group) - What is a Theory or Rationale Behind This?

In Sickness and in Health ... but Not in Debt

Couples spend $30,000 on average planning their Big Day ... but the more you spend, the shorter your marriage is.

- Guys, investing between $2,000 and $4,000 on an engagement ring means you’re 1.3 times more likely to get divorced compared with the more frugal fellows who only allocate between $500 and $2,000.

- For both sexes, spending more than $20,000 on the wedding ups the odds of divorce by 3.5 times compared with couples who keep it between $5,000 and $10,000.