

# DS Visualisation and Analysis

Abbey Waldron

# Designing Experiments

1. Show last week's work to the class
2. How to design experiments
3. Whole group design one series of experiments
4. This week's group assignments

# UK 2011 Census

I want to know the distribution (make a plot) of gender imbalance per local authority area in England and Wales (you can consider all ages together). Tell me the mean, the mode and the error on the mean and mode. Are any regions very unusual? Tip: find 2011 UK census data on [ons.gov.uk](http://ons.gov.uk)!

# Random Student Generator

# Data Sets

PP02 (male usual resident population)

PP03 (female usual resident population)

Use all ages, how to deal with Excel? Export to .csv!

# Define Gender Imbalance

$$\frac{\text{male} - \text{female}}{\text{male} + \text{female}}$$

# Experiments

Up until this point we have been talking about data sets that already exist. What if the data hasn't been collected yet? Designing experiments also forms an important part of the data scientist's skill set.

# Hypothesis: Ask a good question

Hopefully interesting, but more importantly it should be well defined before you plan the experiment. That doesn't mean to say that you can't discover answers to other interesting questions during the course of your experiment! You should however have a goal so that you can design it well.



# Experiments without People

Many areas of science. What about in business?

- ▶ Process/production optimisation
- ▶ Design of new products
- ▶ ...

# Experiments with People

Now you have two problems.

The first decision that you need to make is whether the test subjects will be conscious or unconscious of their involvement in your study.

# Medical Experiments

In medical clinical trials it is impossible (or unethical) to conduct a trial without a subject's knowledge.

The solution employed to get around this is to give some people a placebo (something that does nothing but looks the same as the trial medicine) and to not tell the participants which of them has the real medicine.

# Some Early Medical Experiments I

James Lind (sailors with scurvy)

- ▶ Cider
- ▶ sulphuric acid
- ▶ half-pint of seawater
- ▶ garlic, mustard and horseradish mixture
- ▶ vinegar
- ▶ two oranges and a lemon

# Some Early Medical Experiments II

Edward Jenner (cowpox/smallpox)

# Experiments with Conscious People

In other situations it may be preferable to get information from people with their knowledge (e.g. census data, exit poll surveys) as you can get more information this way.

Of course there is also the problem that people either may not be entirely truthful with you or may not be conscious of their own behaviours.

# Experiments with Concious People

A good example is the optimisation of online web stores, what you really want is for customers to spend more.

You should measure this rather than their opinion about what they like, it could be the most annoying features lead to increased spending!

# Experiments with Unconscious People

In the optimisation of websites the most common technique employed is called A/B testing - some visitors are shown the current version of the website as a control, and some visitors are shown a version with some modification to see if their desired behaviours improve.

Of course you need to decide what you consider a "desired behaviour" and set up tracking to measure them.

Who should you show the new version to?



# Conducting A/B Tests

Use cookies or some other persistent method, otherwise people may see different versions of the website each time they visit and figure out what you're up to.

# Obama Campaign

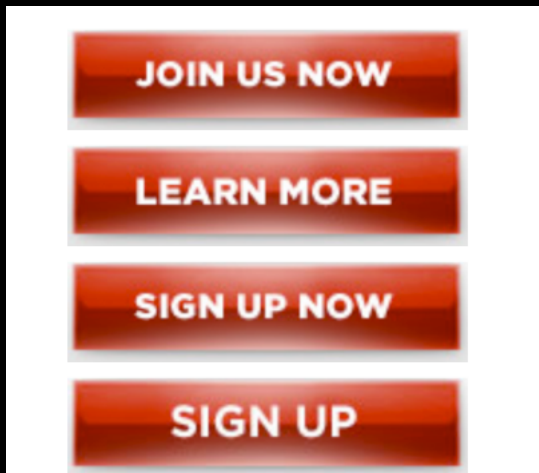
Used A/B testing extensively to increase donations (by around \$60 million)

# Obama Campaign



Optimizely

# Obama Campaign



Optimizely

# Obama Campaign

Combinations (24)		Page Sections (2)		Download:  XML  CSV  TSV    Print		
Relevance Rating	Variation	Est. conv. rate	Chance to Beat Orig.	Observed Improvement	Conv./Visitors	
<b>Button</b> <div>5 / 5</div> <div><div></div><div></div><div></div><div></div><div></div></div>	Original	7.51% ± 0.2%	—	—	5851 / 77858	
	Learn More	8.91% ± 0.2%	100%	18.6%	6927 / 77729	
	Join Us Now	7.62% ± 0.2%	73.5%	1.37%	5915 / 77644	
	Sign Up Now	7.34% ± 0.2%	13.7%	-2.38%	5660 / 77151	
<b>Media</b> <div>5 / 5</div> <div><div></div><div></div><div></div><div></div><div></div></div>	Original	8.54% ± 0.2%	—	—	4425 / 51794	
	Family Image	9.66% ± 0.2%	100%	13.1%	4996 / 51696	
	Change Image	8.87% ± 0.2%	92.2%	3.85%	4595 / 51790	
	Barack's Video	7.76% ± 0.2%	0.04%	-9.14%	3992 / 51427	
	Sam's Video	6.29% ± 0.2%	0.00%	-26.4%	3261 / 51864	
	Springfield Video	5.95% ± 0.2%	0.00%	-30.3%	3084 / 51811	

Optimizely

# Obama Campaign



A mockup of the Obama '08 campaign website. At the top is the Obama '08 logo, consisting of a blue circle with a red and white wave-like design, and the text "OBAMA'08" below it. The main headline reads "CHANGE" in large blue letters, with "WE CAN BELIEVE IN" in smaller blue letters underneath. Below the text is a black and white photograph of Barack and Michelle Obama with their two children. At the bottom of the main content area, there is a "JOIN THE MOVEMENT" link on the left, two input fields for "Email Address" and "Zip Code" in the center, and a red "LEARN MORE" button on the right. The footer is a solid blue bar containing the text "PAID FOR BY OBAMA FOR AMERICA" on the left, a small Obama logo in the center, and the text "CONTINUE to WEBSITE" on the right.

OBAMA'08

# CHANGE

WE CAN BELIEVE IN

[JOIN THE MOVEMENT](#)

[LEARN MORE](#)

PAID FOR BY OBAMA FOR AMERICA

[CONTINUE to WEBSITE](#)

Optimizely

# WARNING

It is ILLEGAL to conduct psychological research without first getting approval from the appropriate country's medical ethics council.

What does this mean for us?

- ▶ Think very carefully about whether your research is telling you about people or people's response to products
- ▶ Avoid deliberate manipulation of emotions that goes beyond that generally used in marketing
- ▶ Don't be evil...

# WARNING II

Take privacy seriously. Internet data leaks can lead to real deaths and prosecutions.



# Class Exercise

# Let's Design One Together

You want to reduce (car) traffic through Rotterdam city center.

# Designing Experiments

I find it is helpful to think about a number of different questions, let's go through some...

# Let's Design One Together

You want to reduce (car) traffic through Rotterdam city center.

- ▶ Why?
- ▶ Who?
- ▶ When?

# Let's Design One Together

You want to reduce (car) traffic through Rotterdam city center.

- ▶ What is the null hypothesis for future experiments?
- ▶ How do you conduct the experiments? Many possibilities?

# Let's Design One Together

You want to reduce (car) traffic through Rotterdam city center.

- ▶ What ideas do you have?
- ▶ How can you/Can you test them?

# Let's Design One Together

You want to reduce (car) traffic through Rotterdam city center.

- ▶ Are you worried about anything, calibration?

# Let's Design One Together

You want to reduce (car) traffic through Rotterdam city center.

- ▶ Is your data collection ethical?
  - ▶ Instead try asking could I think of an evil use for the data - prevent those!



# Week 6 Problems

Next week you are going to give a presentation (10+5 mins, with proper slides) to me and the class about experiments you are going to conduct. I want a serious presentation, like I am your client and you are proposing a new research experiment.

**You may work in groups of up to 5 this week.** I will give each group a different experiment to plan during the class. If one person presents per group that is okay, but I will be asking questions afterwards and everyone should be able to answer them!

# What to think about?

The aim:

- ▶ What is the goal of the experiment?
- ▶ What is your hypothesis? What is your null hypothesis?

# What to think about?

The design:

- ▶ What is the general design of your experiment?
- ▶ Data collection: how will you do it practically?
- ▶ How much data do you think you'll need? Over what time scale?
- ▶ Can you use historical data to help?
- ▶ Is there another experiment you could do to cross check?

# What to think about?

## Ethics:

- ▶ What ethical considerations do/should you have?
- ▶ If you work with people are they conscious or unconscious of what you are doing?
- ▶ Is there anything you are worried about?
- ▶ Consent and privacy: how will you get consent, if necessary?
- ▶ Is your experiment legal?

# What should be in your presentation?

1. A description of the problem you have been given (everyone has different tasks so they won't know!)
2. Examples of real life situations/companies where this may occur
3. Description of your experiment, with a detailed timeline and plan
4. Discussion of difficulties you expect, ethical considerations etc. any of the above questions that yielded interesting answers

# Backup Slides