Asking Questions with the Upworthy Archive

COMM 4940

The Internet

J. Nathan Matias

@natematias

citizensandtech.org

natematias.com

Mariane Aubin Le Quere

PhD Student

Information Science

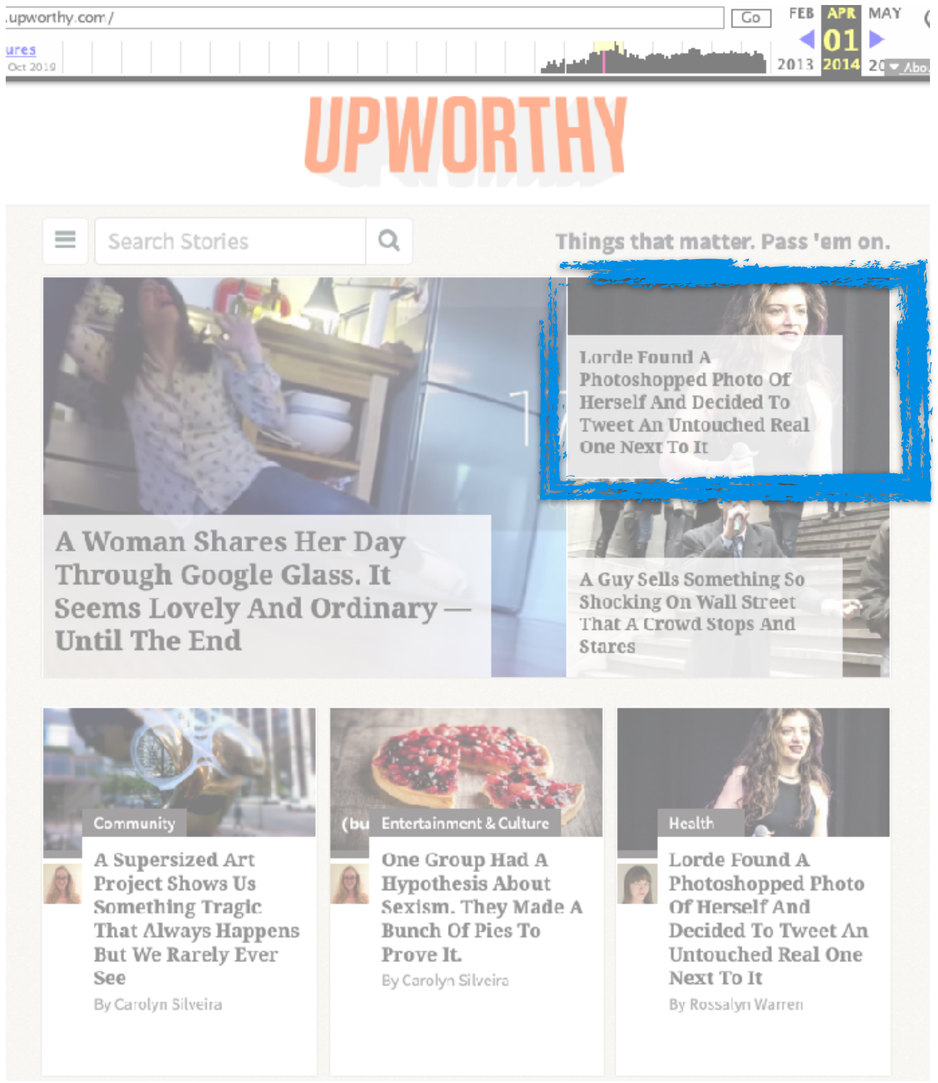
Cornell CALS

Data in the Upworthy Archive

Packages
An article preview link tested by Upworthy

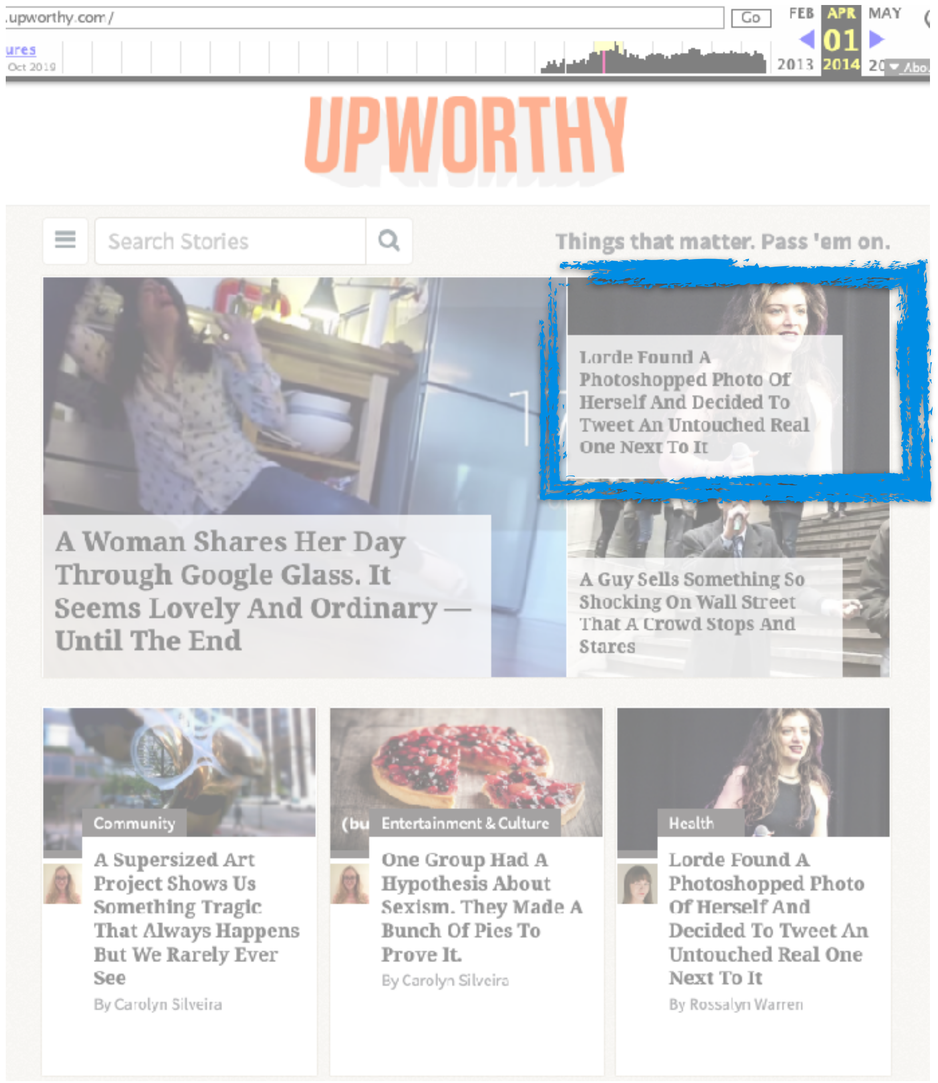
Tests
A group of packages that Upworthy tested alongside each other

Understanding Packages



- **created_at**: time the package was created
- **test_week**: week the package was created
- **clickability_test_id**: test the package was in
- **headline**
- **eyecatcher_id**: image ID (we don't have images)
- **excerpt**: article excerpt (not shown in tests)
- **slug**: internal name (not shown in tests)
- **share_text**: social media description (not shown in tests)
- **impressions**: # who viewed the package
- **clicks**: # who clicked the package

Understanding Packages



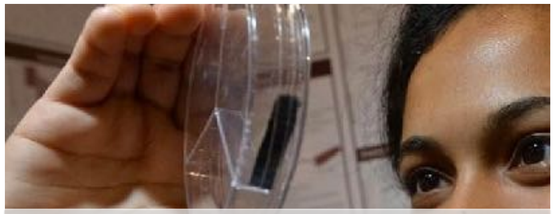
More general terms for a thing being tested (packages was Upworthy's internal term):

- **Conditions** (refers to the experience that the participant is randomly assigned to)
- **Arms** (often synonymous with conditions, especially in studies with more than 2 arms) (**we will use this term in the class**)
- **Interventions** (often used to refer to the act of intervening)
- **Stimuli** (refers to the content of the packages / interventions)

Tests: More than One Package

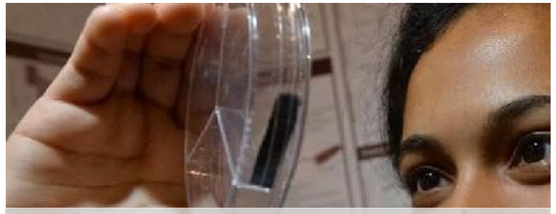
In the Upworthy Archive, tests can compare any number of packages, though most tests compare four packages

A



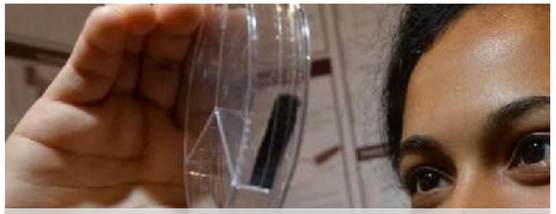
She's Not Just Destined For Greatness, She's Destined To Do Great Things For Women

B



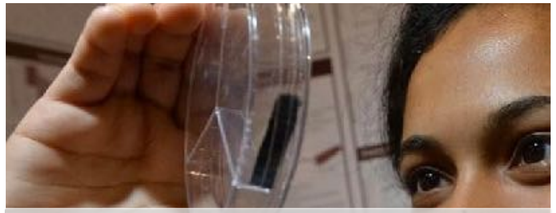
This Young Woman Just Took Silicon Valley By Storm And She's Not Stopping There

C



Feminism 101: This Girl Is Going Places And She's Taking Other Girls With Her

D

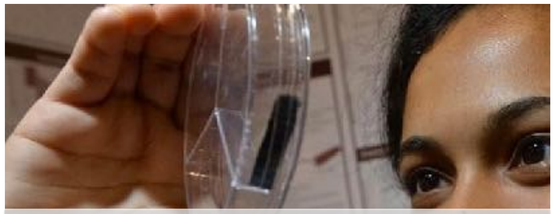


Remember When Math Was “Too Hard” For the Ladies? Not So Much.

Tests Vary More than Headlines

Since images vary and some differences between packages can be caused by image differences, not all headlines are comparable

A



She's Not Just Destined For Greatness, She's Destined To Do Great Things For Women

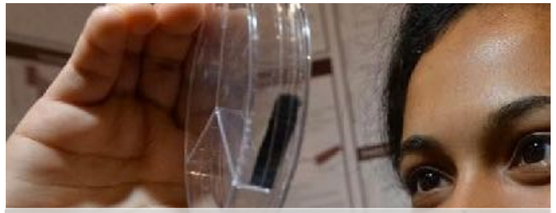
B

This Young Woman Just Took Silicon Valley By Storm And She's Not Stopping There

C

Feminism 101: This Girl Is Going Places And She's Taking Other Girls With Her

D



Remember When Math Was "Too Hard" For the Ladies? Not So Much.

Hypothetical: we don't have the images, just the image IDs

Protecting Ourselves from P-Hacking

- **Confirmatory research:** In experiments, we define analysis **before** collecting data. The confirmatory process protects us from p-hacking, since we have few or no adjustments after seeing the data
- **Exploratory research:** In research with an already-collected dataset, standard practice is to keep testing ideas until we are confident about the answer, putting us at risk of p-hacking

Cross-Validation: A Confirmatory Process for Exploratory Analysis

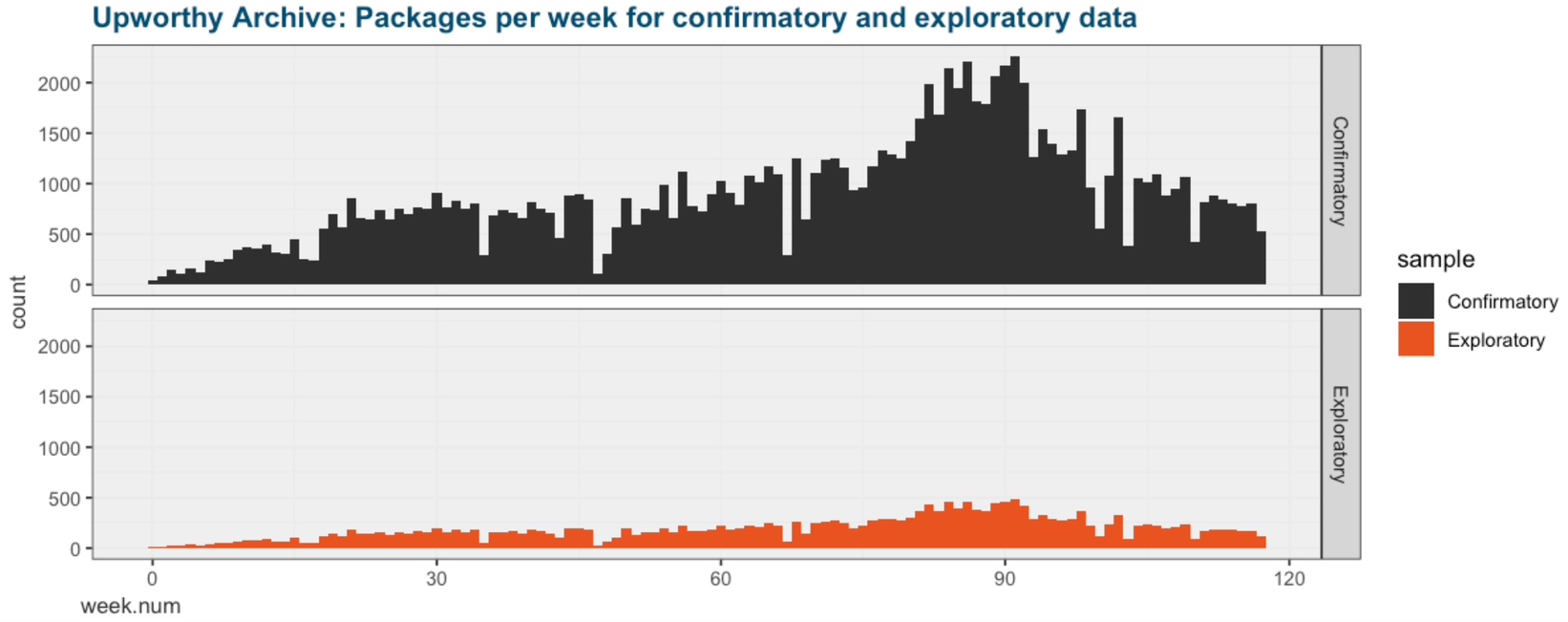
- **Cross-validation** involves testing a statistical model on multiple samples:
  - Researchers analyze **exploratory data** to choose a final model
  - The model is run on **confirmatory data**, and these results are accepted as final

Upworthy Archive Cross-Validation

- **Exploratory data**: students will work with a dataset of 4,873 tests & 22,666 packages
- **Exploratory analysis**: teams will use this dataset to develop a proposed model & submit code
- **Confirmatory data**: student models will be tested on a dataset of 22,743 tests with 105,551 packages
- **Confirmatory analysis**: conducted by the professor with results shared back to student teams

Upworthy Archive Datasets

The exploratory dataset is a random sample of all tests within each week in the archive. This makes it possible to test ideas about changes over time



How to ask questions with the archive

1. Identify a **question / theory to test**
2. Develop **a method to select relevant & comparable packages** & **tests**
3. **Meta-analyze** the results

What is a good question?

1. It contributes to an important **practical** or **scientific** conversation (or both)
2. The archive includes data that can answer the question
3. By the end of the semester, your team will have the skills to ask the question

Question example: well-known people

Does **including a famous person's name** in a headline **increase the number of clicks** it receives?

Checking if your question is askable

1. Can you detect headlines **in a way that makes sense**? (how would you identify notable people?)
2. Are there **cases in the data**?
3. Do those cases **appear in the same tests**?

Identifying Notable People

1. Collect lists of notable people from TIME Magazine and IMDB
2. Create a method to identify headlines that **include at least one notable person's name**
3. Compare packages within tests to confirm:
   - at least one headline has a notable person's name
   - one headline doesn't have any of those names

Valid Tests Need a Comparison

Headlines in "**54eb434831386500264200000**" mention two notable people, but it's **not a valid test** because **all headlines include the same names**

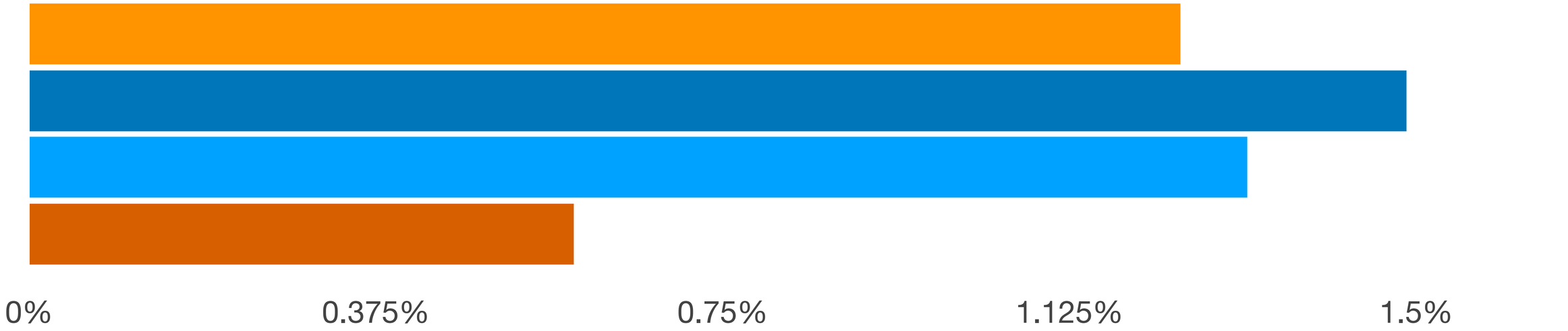
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later
- Beyoncé Had A Moment With Ed McMahon That Fueled Her Success, But It Also Hit Her Differently Later

Valid Tests Need a Comparison

Packages in “**54eb434831386500264200000**” have headlines that **do mention** notable people and headlines that **don’t mention them**

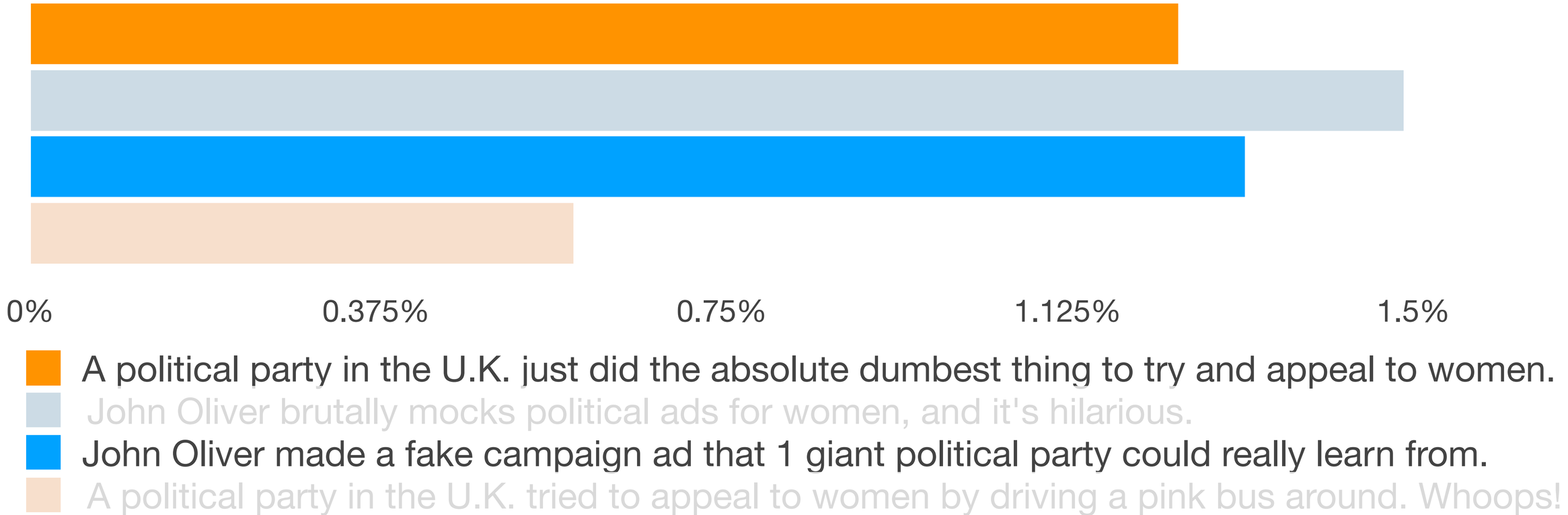
- **A political party in the U.K. just did the absolute dumbest thing to try and appeal to women.**
- **John Oliver brutally mocks political ads for women, and it's hilarious.**
- **John Oliver made a fake campaign ad that 1 giant political party could really learn from.**
- **A political party in the U.K. tried to appeal to women by driving a pink bus around. Whoops!**

Choosing Packages to Compare



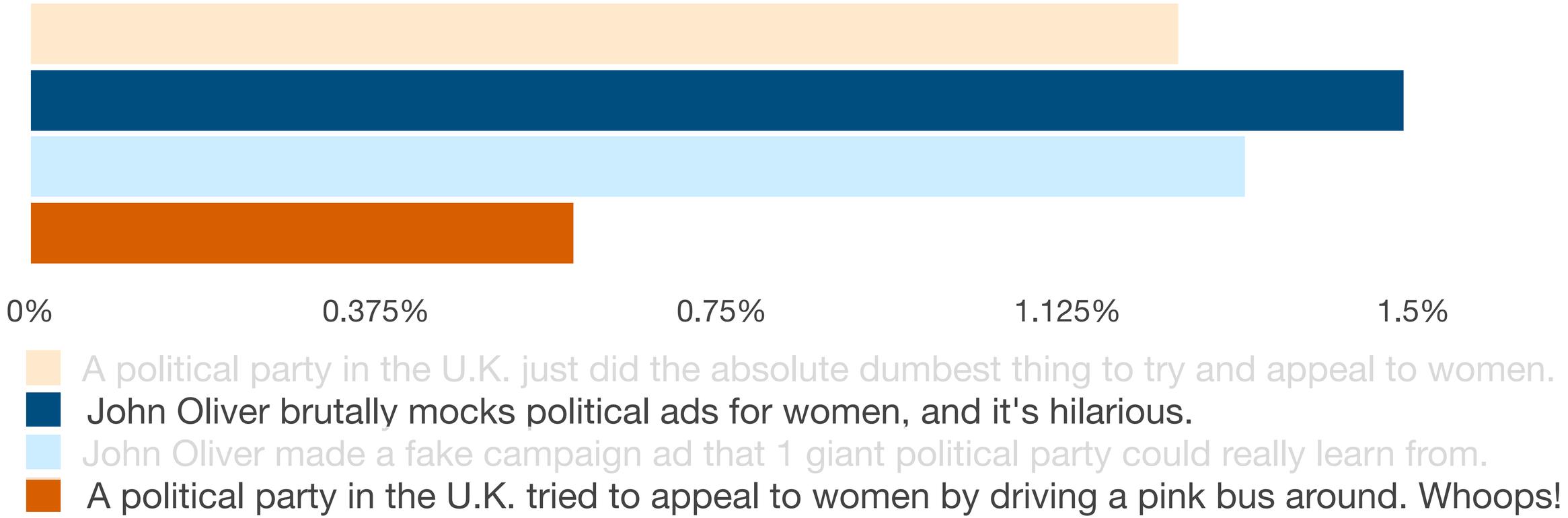
Choosing Packages to Compare

If you choose these two, you might find **small/no** effect.



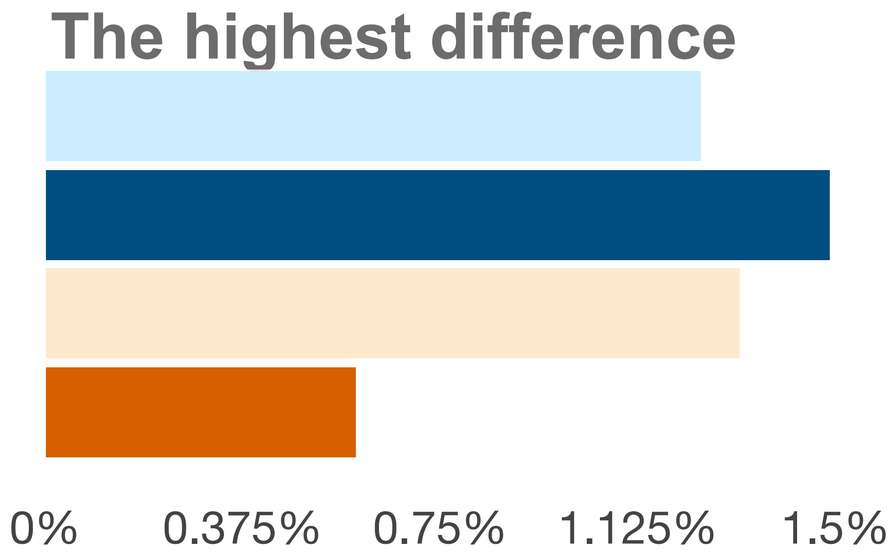
Choosing Packages to Compare

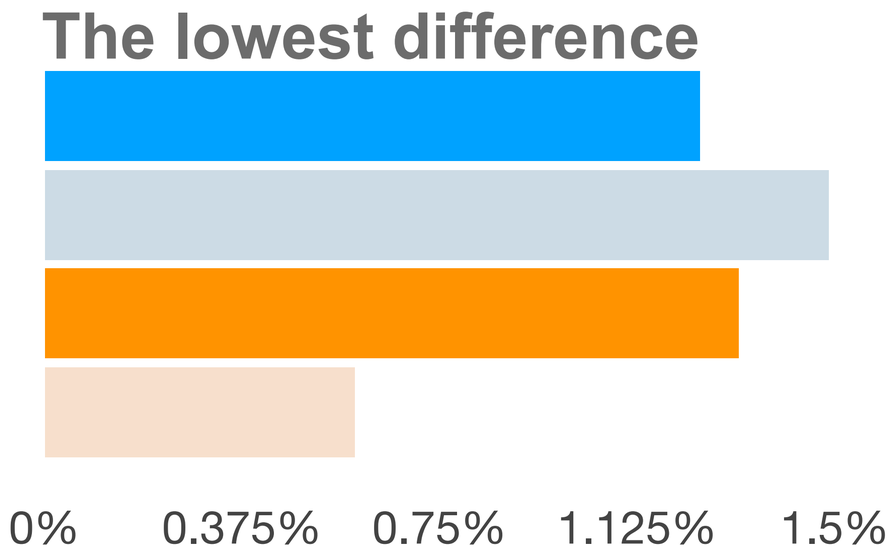
If you choose these two, you might find a **large** effect.



Solution: Upper & Lower Bounds

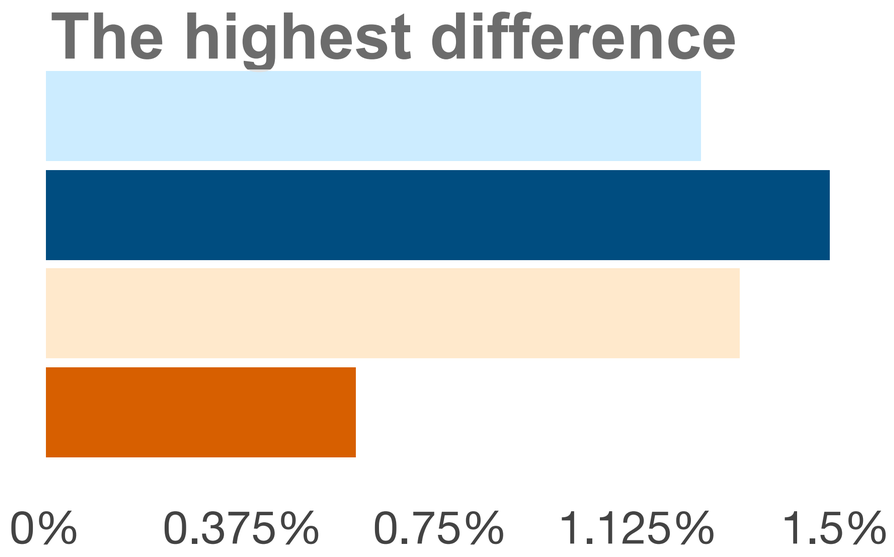
Create **two datasets**: one to identify the highest possible effect (largest difference) & another to identify the lowest possible effect (lowest difference).

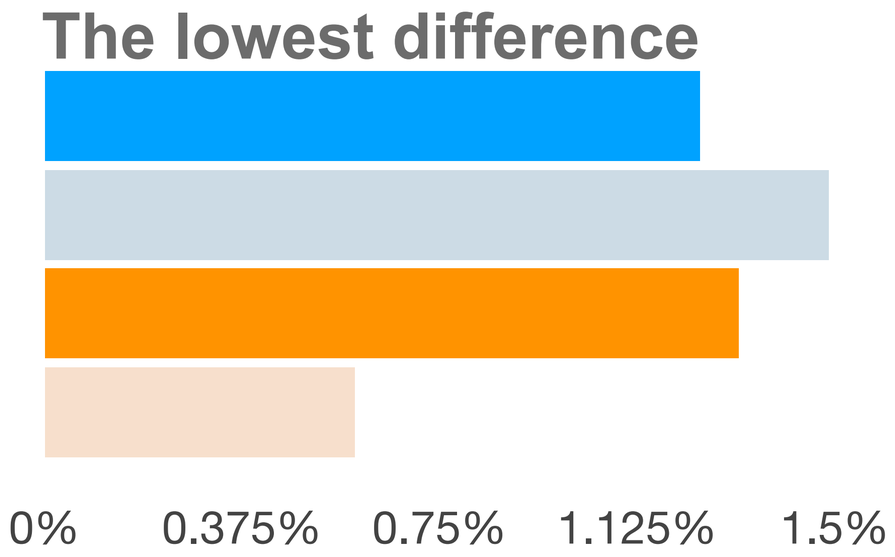




Solution: Upper & Lower Bounds

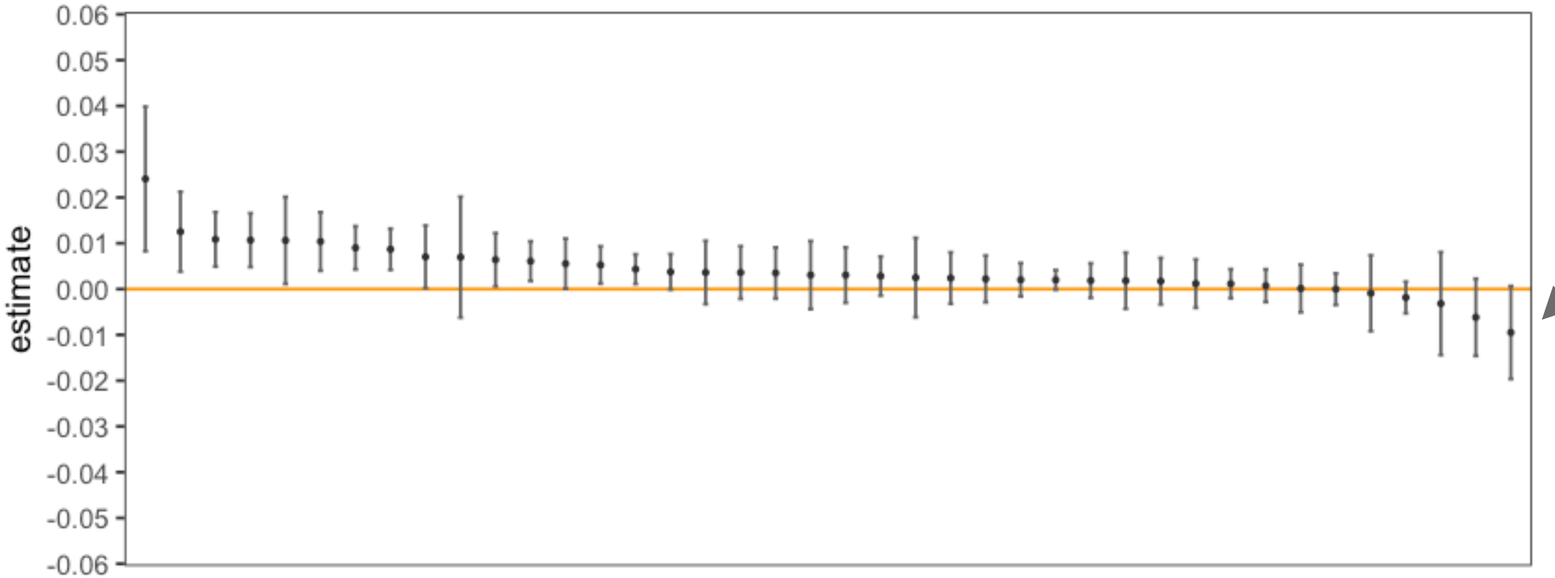
In this test, adding a notable person's name has an effect between **.07 of a percentage point (lower)** and **0.9 of a percentage point (upper)**





Analyzing Multiple Tests Together

Once we have selected tests, we can analyze the combined effect across many tests. The practice of combining studies is called “meta-analysis.”



Headline A/B tests conducted by Upworthy that included a notable person's name versus a headline without a notable person's name (n=40 tests and 270,177 viewers)

Each of these is the effect size of including a notable person’s name in the headline

Example Code

Full examples for working through the unworthy archive project data are available at:

github.com/natematias/design-governance-experiments/tree/master/assignments/upworthy-archive-project