Plan:

1. Define Data Science
2. Review a few data science examples

Intro to Data Science

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu

Why this course?

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

Harvard Business Review

50 Best Jobs in America

Awards

- Best Places to Work
- Top CEOs
- Best Places to Interview

Lists

- **Best Jobs**
- Best Cities for Jobs
- Highest Paying Jobs
- Oddball Interview Questions

This report ranks jobs according to each job's Glassdoor Job Score, determined by combining three factors: number of job openings, salary, and overall job satisfaction rating.

United States | 2018

0 Shares

1 Data Scientist

4.8 / 5
Job Score

4.2 / 5
Job Satisfaction

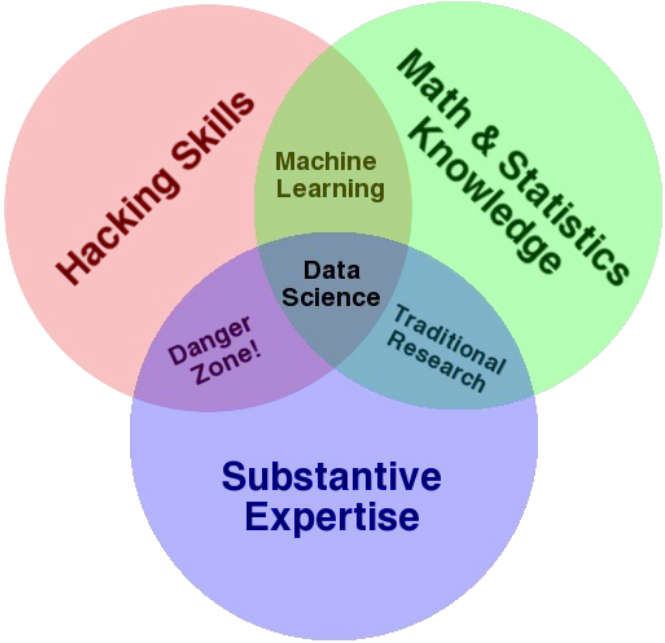
$110,000
Median Base Salary

4,524
Job Openings

View Jobs

https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

What is data science?



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Defining Data Science

a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data.[3] It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science. -Wikipedia

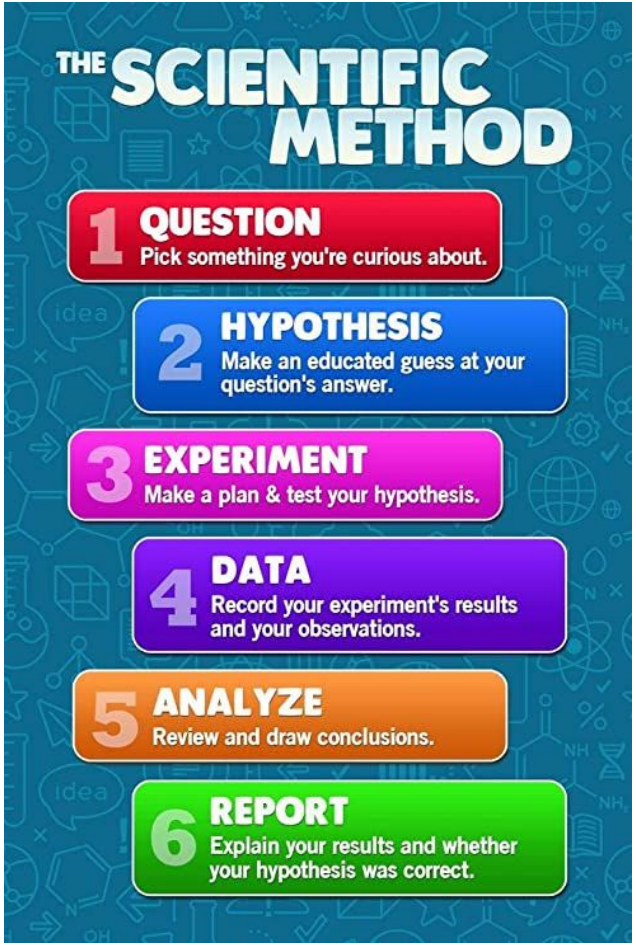
"This coupling of scientific discovery and practice involves the collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and interdisciplinary actions." -David Donoho ("*50 years of Data Science*

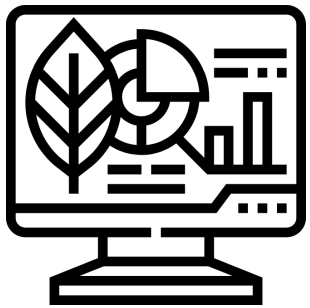
"an emerging discipline that draws upon knowledge in statistical methodology and computer science to create impactful predictions and insights for a wide range of traditional scholarly fields" - from a panel Rafael Irizarry moderated, shared on SimplyStatistics ("*The role of academia in data science education*")

"an umbrella term used by organizations to describe the processes used to extract value from data" -Rafael Irizarry's personal definition in "*The role of academia in data science education*"

"The study of how the quantification of observable phenomena can lead to human understanding of the processes giving rise to those phenomena—or even the ability to predict future outcomes absent human understanding—and why certain phenomena require more or less data to lead to human understanding and/or prediction accuracy". -Brad Voytek's definition

"The scientific process of extracting value from data"


THE SCIENTIFIC METHOD
1 QUESTION
Pick something you're curious about.
2 HYPOTHESIS
Make an educated guess at your question's answer.
3 EXPERIMENT
Make a plan & test your hypothesis.
4 DATA
Record your experiment's results and your observations.
5 ANALYZE
Review and draw conclusions.
6 REPORT
Explain your results and whether your hypothesis was correct.



Data scientists ask interesting questions & answer them with data

A silly example:
What questions should I ask on a first date?

The Best Questions For A First Date

How asking certain questions can reveal much more

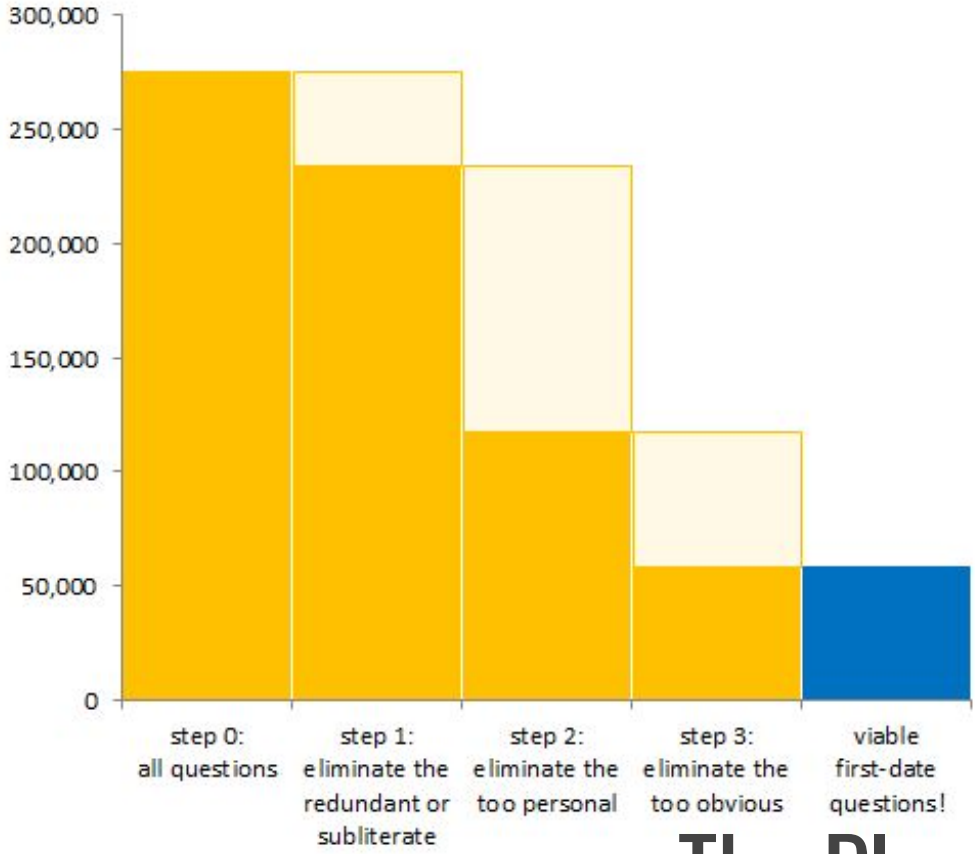
OkCupid Follow

Apr 19, 2011 · 5 min read

This post is our attempt to end the mystery. We took OkCupid's database of 275,294 match questions—probably the biggest collection of relationship concerns on earth—and the 776 *million* answers people have given us, and we asked:

> *What questions are **easy to bring up**, yet correlate to the deeper, unspeakable, issues people actually care about?*

https://theblog.okcupid.com/the-best-questions-for-a-first-date-dba6adaa9df2



The Plan

https://theblog.okcupid.com/the-best-questions-for-a-first-date-dba6adaa9df2

If you want to know…

Do my date and I have long-term potential?

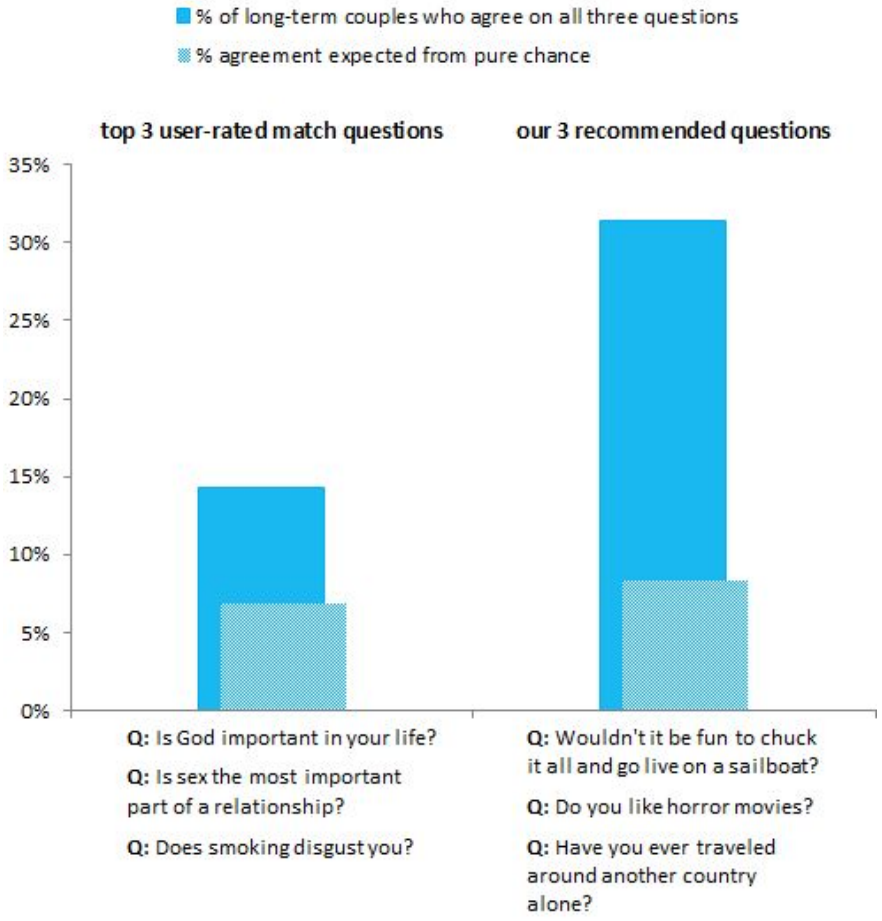
Ask your date (and yourself!)…

Do you like horror movies?

Have you ever traveled around another country alone?

Wouldn't it be fun to chuck it all and go live on a sailboat?

https://theblog.okcupid.com/the-best-questions-for-a-first-date-dba6adaa9df2



https://theblog.okcupid.com/the-best-questions-for-a-first-date-dba6adaa9df2

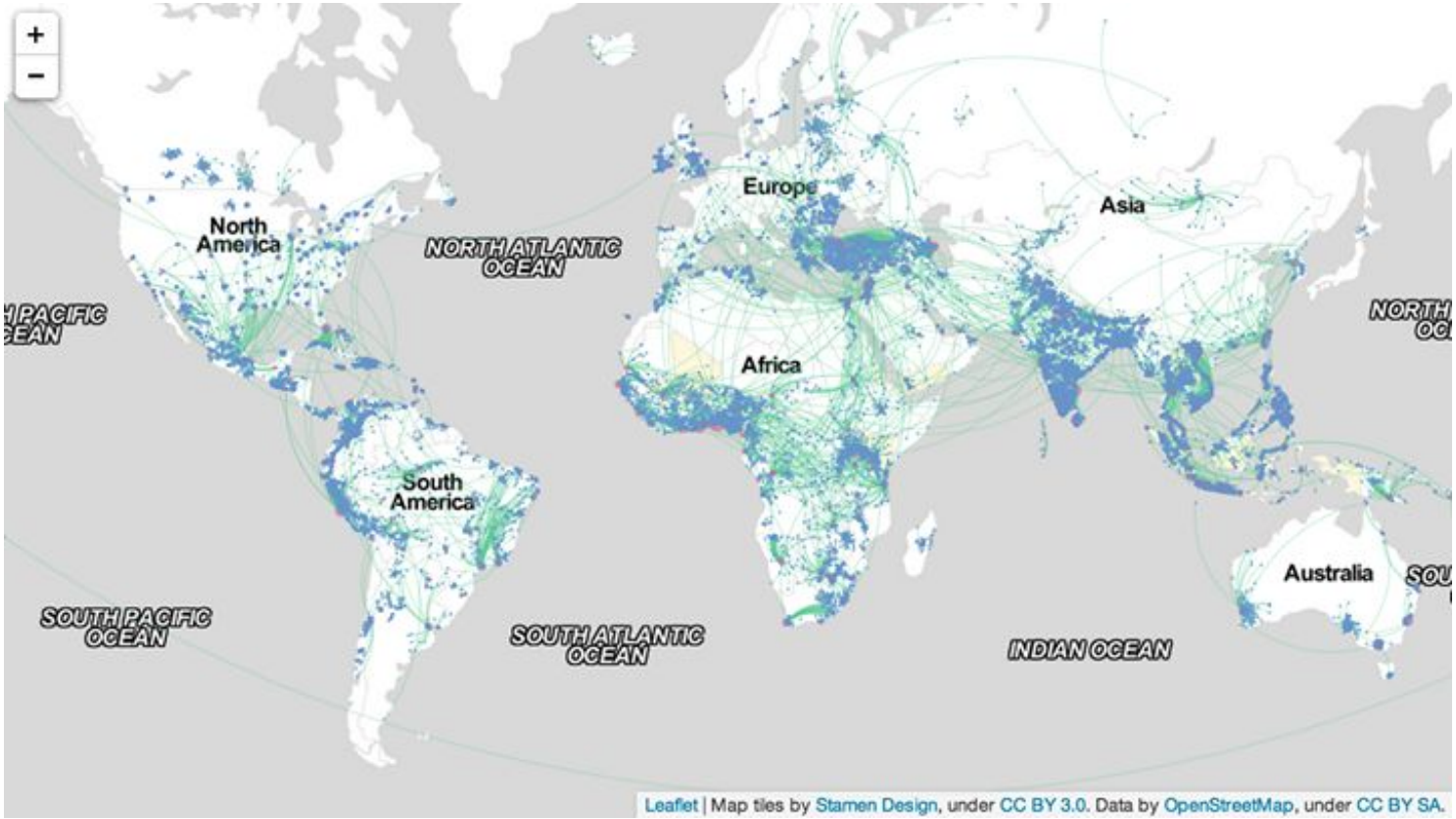
A clever example:
Studying coordinated migration

"A flow of population from city A (*hometown*) to another city B (*current city*) is **considered a coordinated migration if**, among the cities in which people from hometown A currently live, city B is the city with the largest number of individuals with current city B, and hometown A."

https://www.facebook.com/notes/facebook-data-science/coordinated-migration/10151930946453859

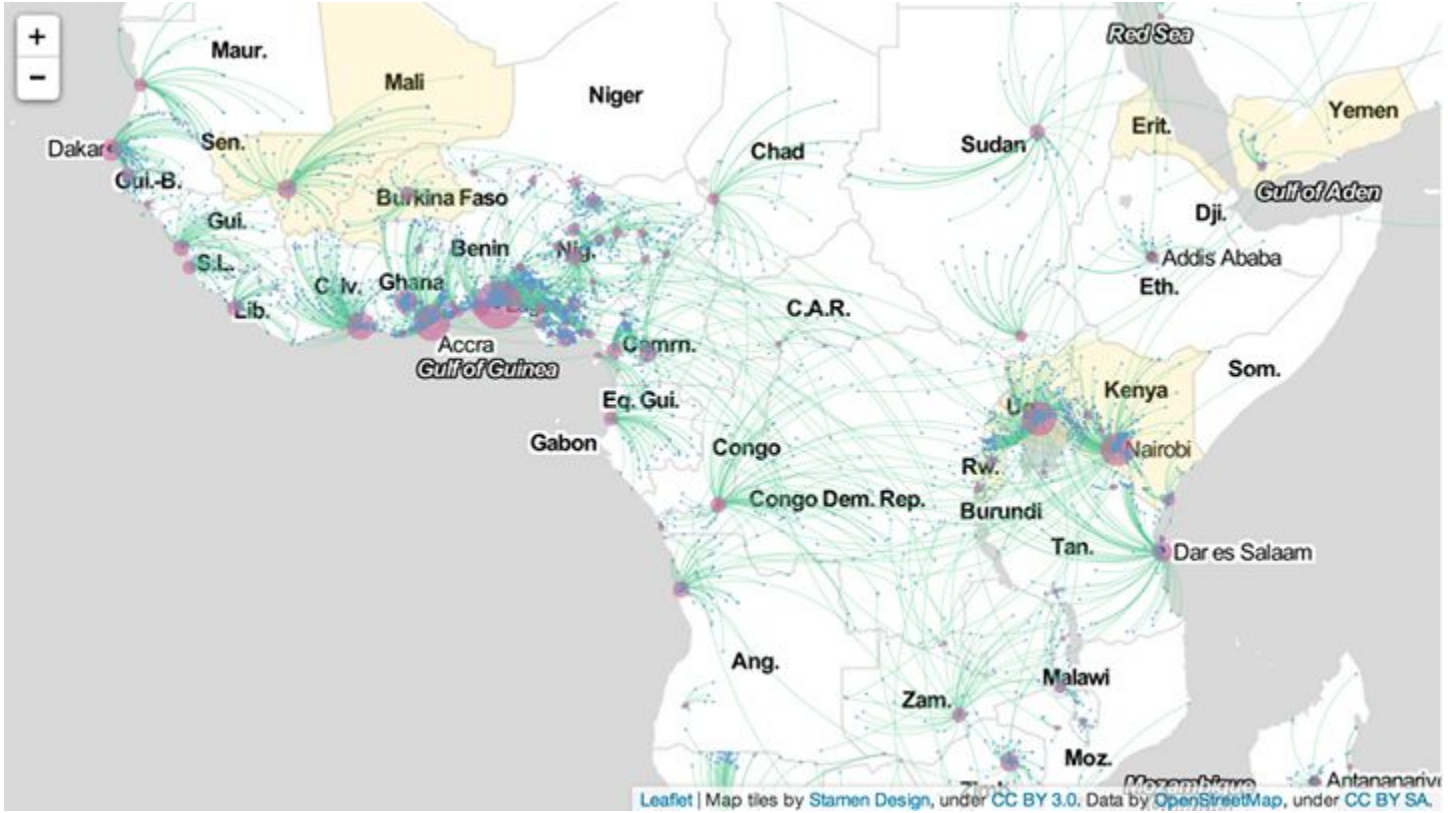
“To study between-city coordinated migration, we examined *aggregate, anonymized data* on all users who list both their hometown and their current city on their Facebook profile.”

https://www.facebook.com/notes/facebook-data-science/coordinated-migration/10151930946453859



Worldwide Coordinated Migration

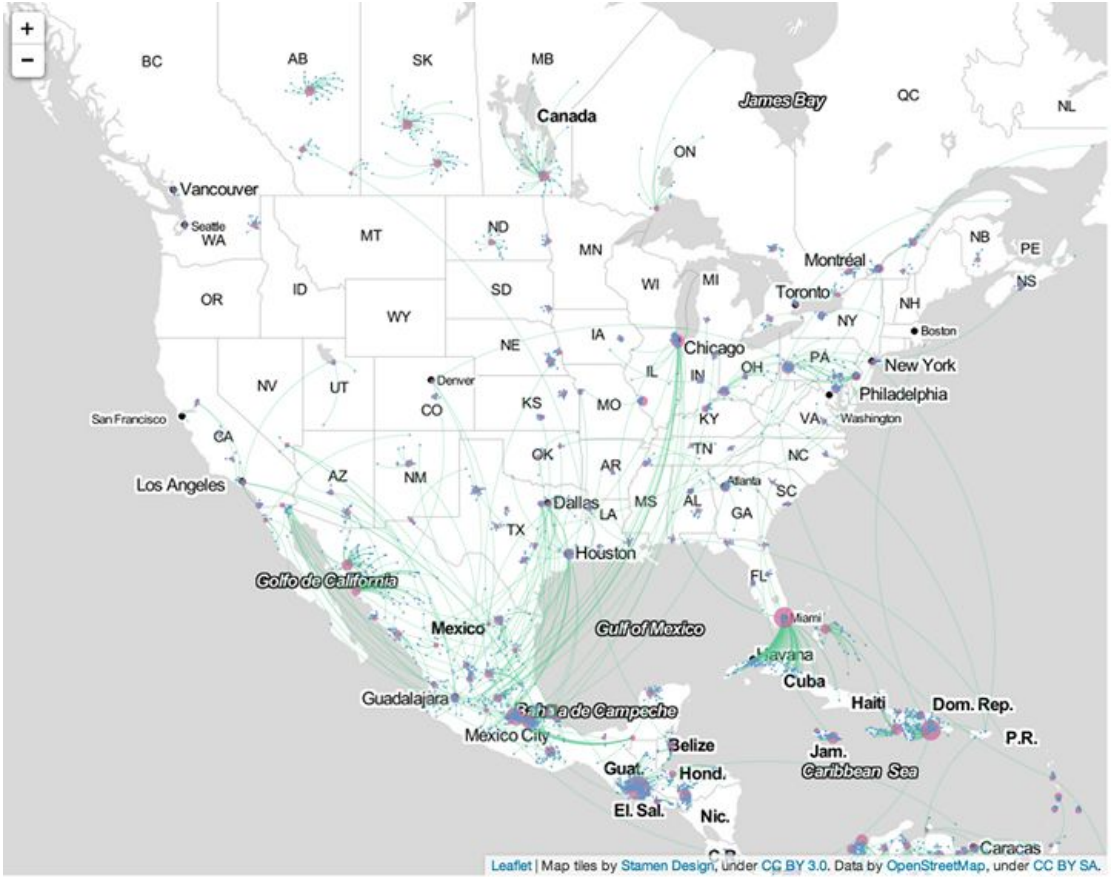
https://www.facebook.com/notes/facebook-data-science/coordinated-migration/10151930946453859



Major destinations of coordinated migration are in rapidly urbanizing countries

https://www.facebook.com/notes/facebook-data-science/coordinated-migration/10151930946453859

Different types of international coordinated migrations have the United States as destination.

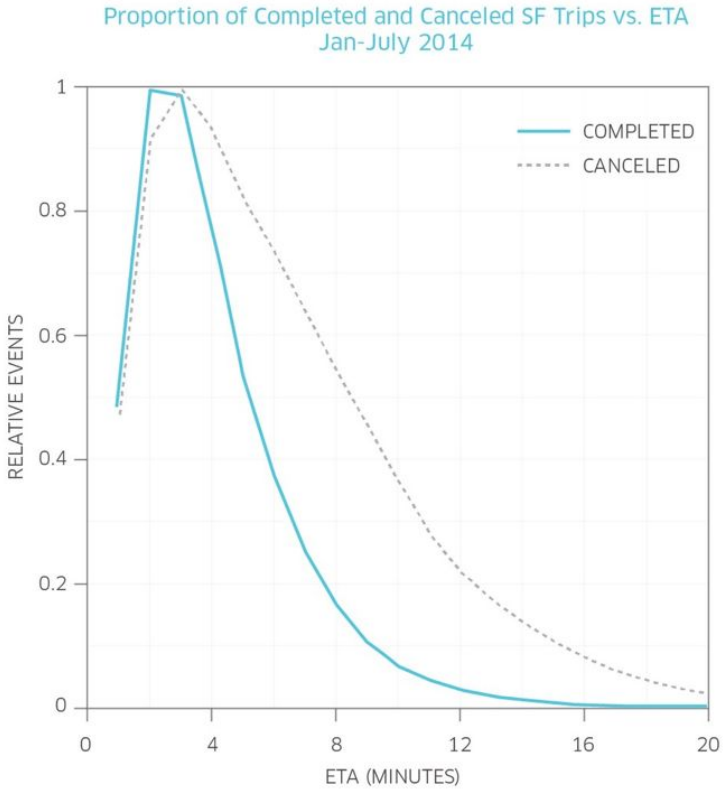


Cuba and Mexico demonstrated highest volume of coordinated migration to the US

https://www.facebook.com/notes/facebook-data-science/coordinated-migration/10151930946453859

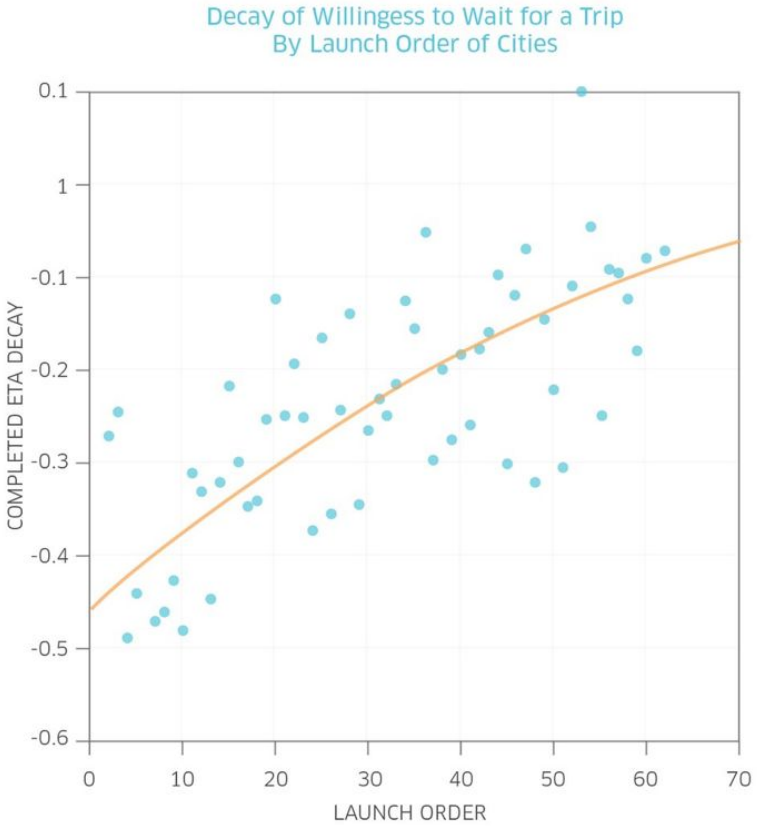
A gig-economy example:
How do I improve ride sharing?

Rider patience



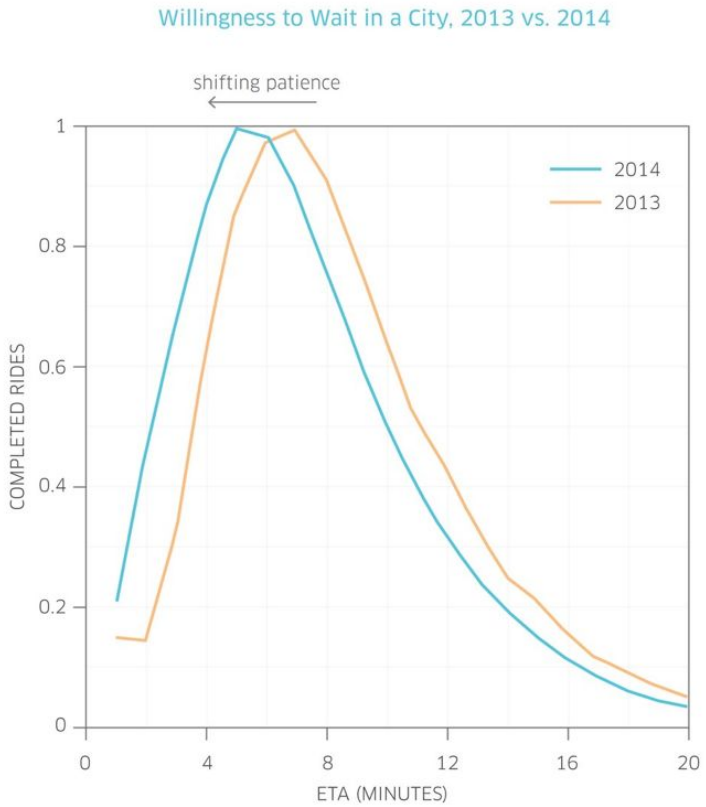
Source: Uber, Inc. (Voytek)

Rider patience



Source: Uber, Inc. (Voytek)

Rider patience



Source: Uber, Inc. (Voytek)

track purchases and online interactions

identify who to focus on during political campaigns

identify patterns in human behavior

build systems that make data more accessible

make governments more effective

determine most effective advertising strategies

improve society (homelessness, crime, access, etc.)

improve system efficiency

ask interesting questions