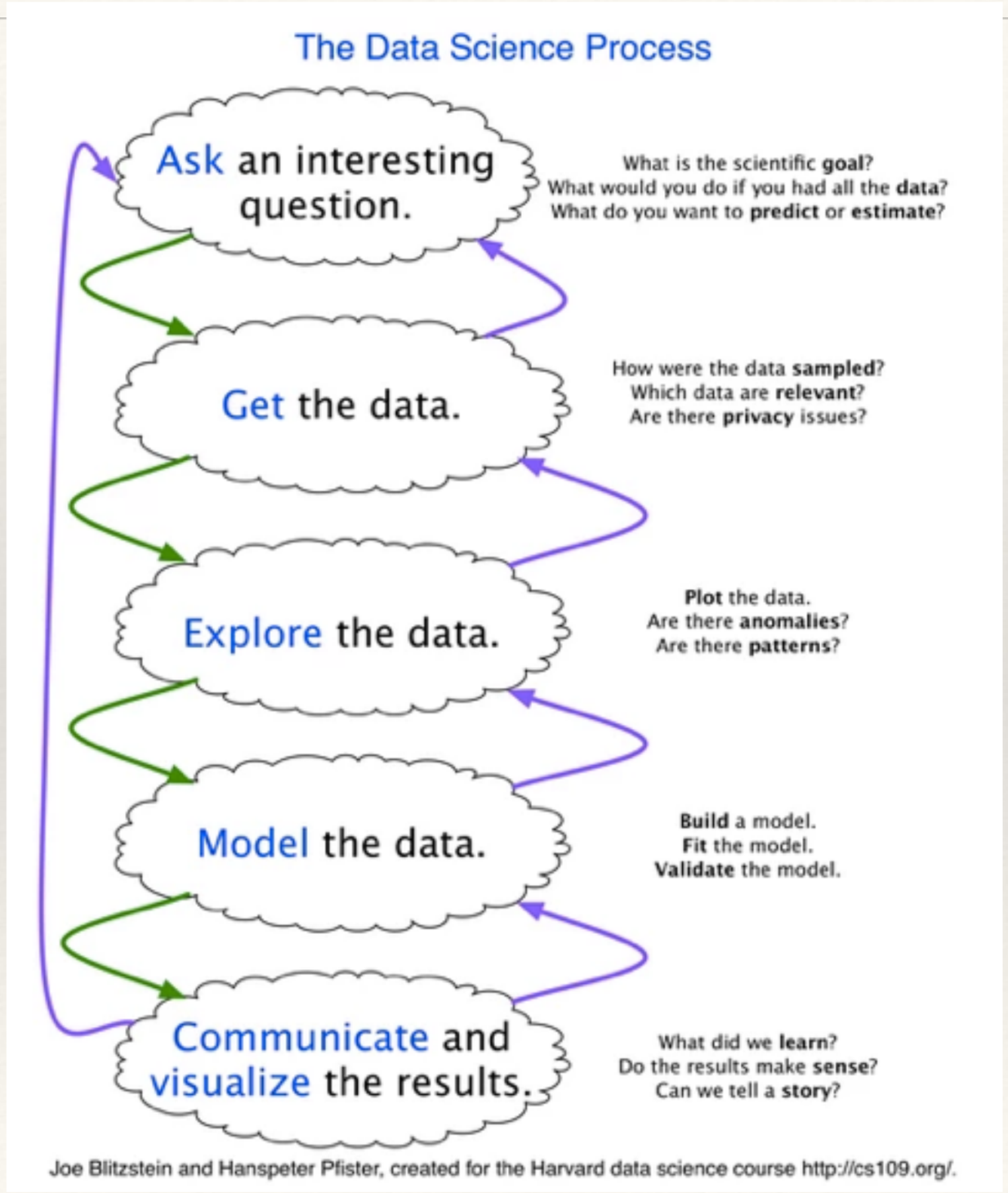
# Design and Process

Knowing what you want to build and why

# Process


The Data Science Process
Ask an interesting question.
What is the scientific goal?
What would you do if you had all the data?
What do you want to predict or estimate?
Get the data.
How were the data sampled?
Which data are relevant?
Are there privacy issues?
Explore the data.
Plot the data.
Are there anomalies?
Are there patterns?
Model the data.
Build a model.
Fit the model.
Validate the model.
Communicate and visualize the results.
What did we learn?
Do the results make sense?
Can we tell a story?
Joe Blitzstein and Hanspeter Pfister, created for the Harvard data science course http://cs109.org/.

# Process


5 LOOM
DATA Storytelling
THE ART AND SCIENCE OF SOCIAL MEDIA METRICS
PACKAGE Never force your data into a package--package your story from the data. The way you apply final data visualization and craft a modern data story needs to be flexible and agile like your strategy. Less data, more story. Less linear, more actionable. Less data showing-off and more zeroing in on opportunity. Modern social media data storytelling is all about adjustments--and a well-packaged data story will guide you to maximum results.
CAPTURE Capture all--report on less. All social data is important data, but not just for reporting--for shaping and informing real-time campaign adjustments. Since social media is about directly communicating with your audiences--understanding what your users tell you via data is a critical part of serving them better.
ADJUSTMENTS
CONTEXT One of the most important and often forgotten layers of storytelling is context. Context isn't just about back-patting as you surpass your competitor in weekly virality. It's about answering the question "so what?" Every report should include competitive, industry and aspirational benchmarks but moreover, by applying context early in the process, you may uncover new stories from this layer of data.
ANALYZE Reams of unusable data need some visualization before they can be analyzed. Utilize multiple types of comparative charts, grids, and visuals to help give data a storytelling boost. Don't be afraid to uncover a story that requires you to drill back down and capture new data. The best analysis happens fluidly.

# Process

- OSEMN: Obtain-Scrub-Explore-Model-Interpret
  - Obtain datasets
  - Clean, combine, transform data
  - Explore the data
  - Try models (classification, machine learning etc)
  - Interpret and communicate your results

http://machinelearningmastery.com/how-to-work-through-a-problem-like-a-data-scientist/

# First, ask a good question

- Understand your target audience
- Write hypotheses that can be explored
  - Do people have more phones than toilets?
  - How is Ebola spreading?
  - Is using wood fires sustainable here?
  - Can we feed 9 billion people?

(Simple, Actionable,Incremental)

# Exercise: Getting Started

- Is there a story you want to tell with data?
- Do you have questions that data might help with?
- Is there a dataset that you'd like to explore?

- List these.
  - Questions need to be simple, actionable, incremental
  - Datasets could be anything - tables, images, maps, sensor feeds; anything.

# Know your Audience

- Personas
  - Who are you trying to influence / inform?
- User stories
  - What are their goals?

Human-centred design for data: http://datascopeanalytics.com/what-we-think/2013/10/23/data-human-centered-design

UX training sites: http://theuxreview.co.uk/series/beginners-guide-to-ux/ http://www.uxapprentice.com/

# Personas

- Get to know the people who will use your system
- Understand their problem
- Understand how people already solve that problem
- Create **personas:** examples of each type of user
  - http://theuxreview.co.uk/personas-the-beginners-guide/

# Ushahidi Persona

## Guillermo // News Gatherer


"I work for a large news organisation, and we want to find new ways to source and tell stories. Crowdsourcing helps us get a better understanding of big events as they unfold. Publishing reports from citizens also helps us differentiate ourselves competitively."

### Overview

Guillermo's job is focused on utilising social media for his news organisation. He uses social media to gather information about emerging events.

His goal is both help journalists source new and different stories, and also help connect the outlet better with its audience.

He uses Ushahidi on occasions when there is a big event, such as civil unrest or a natural disaster.

With this focus, he is prepared to invest time in getting to know Ushahidi. While he'd prefer everything to work perfectly right out of the box, he knows that it's important to customise things so it's more effective.

He's not a technical person, and so relies on the IT people at his office a lot to get the software up and running as he needs it. They can be slow sometimes, so he'd rather not depend on them.

### Satisfiers

Getting a deployment up and running quickly.

Making sure the deployment is visually compelling and professional.

Making it easy for citizens to submit reports of all different media types.

Quick and accurate report verification.

Making it easy for journalists to uncover interesting and useful content.

### Frustraters

Quality of reports is often low; poor descriptions or highly opinionated.

Journalists are often not interested in using Ushahidi to help source their stories; they sometimes don't see the value.

### Usage scenarios

Configure deployment to have the right categories, verification schema, visual presentation.

Set up users with different editing permissions, and permissions to see different levels of information.

Define report structure and permissions.

Coordinate with verification and geolocation volunteer team managers to make sure the flow of reports are being processed.

Share sample outputs with management and journalists to help them start using the platform.

Periodically review the reports and outputs to make sure that everything is running correctly.

| | |
|---|---|
| Technical literacy | ●●○○○ |
| Customisation needs | ●●●○○ |

| | |
|---|---|
| Deployment team | 20-30 |
| Reporters | 500-1000 |

| | |
|---|---|
| Report volume | 100 per day |
| Deployment duration | 2 months |

smallsurfaces Ushahidi

# User Stories

- Look like this:
  - **As a** <role>
  - **I want to** <goal>
  - **in order to** <benefit>

- For example:
  - As a minister for agriculture, I want to know where wheat crops are underperforming and why, so I know where to concentrate resources like education
  - As a director of tree services, I want to predict the trees that might become dangerous in storms, so I can send crews out to manage them before that happens

# Exercise

- Think about the people you want to inform / influence
- Write 1-paragraph persona description for 1-2 of these
- Write 1 or more user stories for them:
  - **As a** <role>
  - **I want to** <goal>
  - **in order to** <benefit>

# Think about Ethics

- You're responsible for your data outputs.
- Could your visualisations increase risk to anyone?
- How bad data fed the Ebola epidemic, New York Times
- How will you respect privacy and security?

# Obtain/Scrub

- find data
- get that data (manually or automatically)
- reformat the data
- clean the data

# Exercise!

- List the data you need for your user stories
- Look for that data (see “Places to look for data” directory)
- Think about what you’ll do if data isn’t available
  - Use proxy datasets
  - Create datasets: surveys, crowdsourcing etc
- Download some example data (if available)

# Explore/ Model

- explore data
- model data
- interpret
- predict
- test hypotheses

# Interpret

- Interpret and communicate your data and results
- Results aren't useful if they don't *do* something
  - e.g. Persuade a decision-maker
- Good visualisation = insight, persuasion
- Great visualisation = a compelling story using data

# Exercise!

- Think about how you want to communicate:
  - Which visualisations might be useful
  - What data would users want to drill down into?
- Look at example visualisations - think about what inspires you, or might fit your use cases
  - Tableau gallery: https://www.tableau.com/public/gallery
  - D3 gallery: https://github.com/mbostock/d3/wiki/Gallery