Recap

Overview

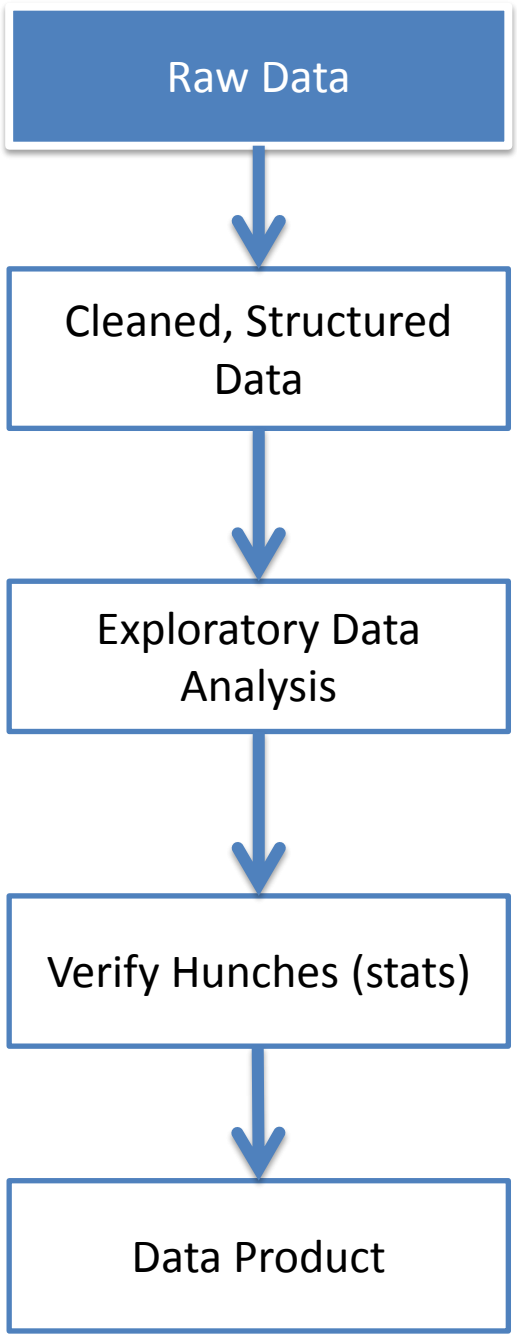
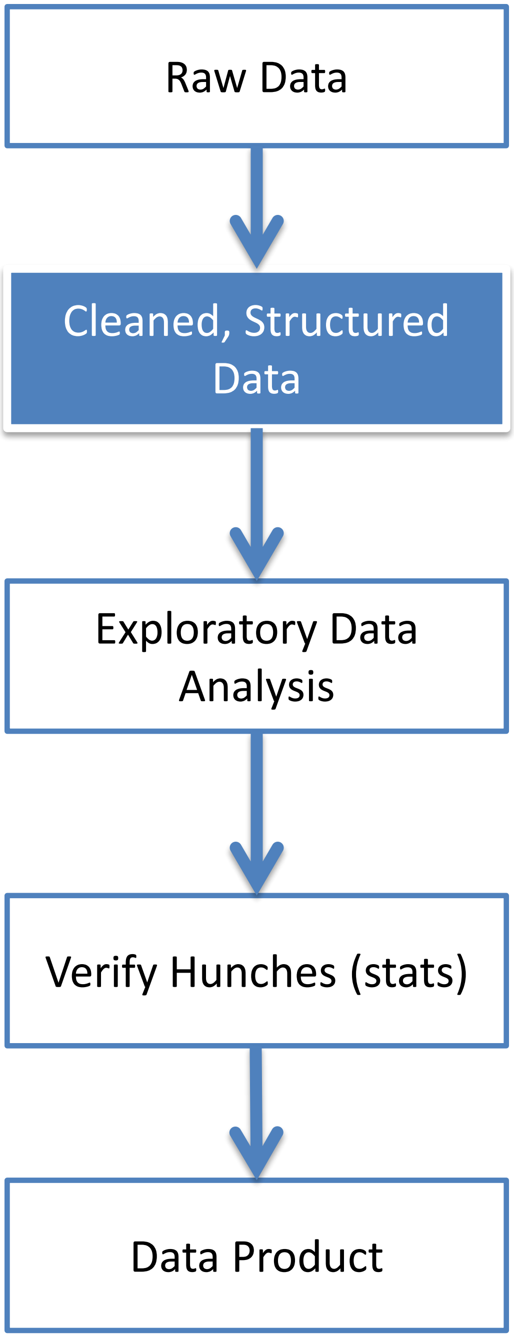
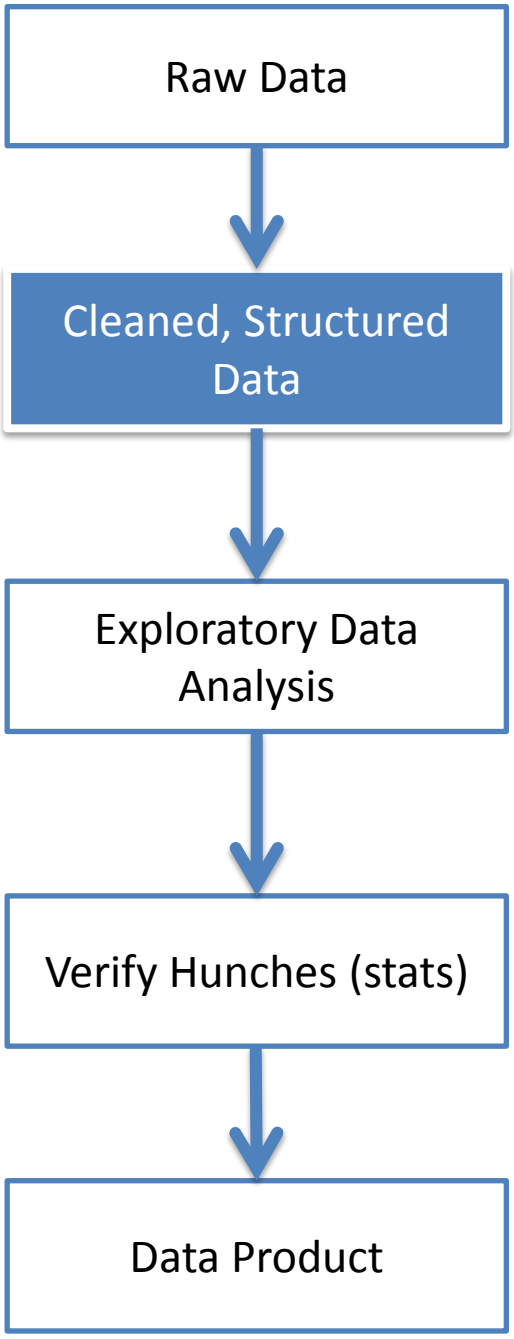


Image of Schedule A-P, showing two contributions to Obama for America. Data includes full name, date of contribution, and contribution amount.



C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-
07,"","","","SA17A",288757C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-
07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757



C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-
07,"","","","SA17A",288757C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-
07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX"."AZ","85051"."RETIRED"."".25,11-JAN-07."".".".","SA17A",288757

"SANTA CLAUS","SNOWMAN",

C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","BROWN, CHARLENE","EAGLE RIVER","AK","99577","","STUDENT",25,01-MAR-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","KELLY, RAY","HUNTSVILLE","AL","35801","ARKTECH","RETIRED",25,25-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CINGEL, KEITH","SEVERN","AL","20999","SANTA CLAUS","SNOWMAN",50,17-MAY-07,"","","","SA17A",305408
C00420224,"P80002983","Cox, John H","DUNAWAY, JONATHON","DEATSVILLE","AL","36022","CSC","TECHNICAL MANAGER",10,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","TERRY, R.S. MR. SR.","SHEFFIELD","AL","35660","RETIRED","",25,18-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","CANADY, DALE","PHOENIX","AZ","85051","RETIRED","",25,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","LORENZ, DWIGHT","SUN CITY","AZ","85351","NONE","RETIRED",20,12-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","STEWART, MICHAEL","CHANDLER","AZ","85224","DYNAMIC ENERGY","TECHNICIAN",5,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","ROSENTHAL, ARNOLD","CAREFREE","AZ","85277","RETIRED","",10,11-JAN-07,"","","","SA17A",288757
C00420224,"P80002983","Cox, John H","VADNAIS, DOROTHY","SAN DIEGO","CA","92116","","RETIRED",10,10-JAN-07,"","","","SA17A",288757

Raw Data

↓

Cleaned, Structured Data

↓

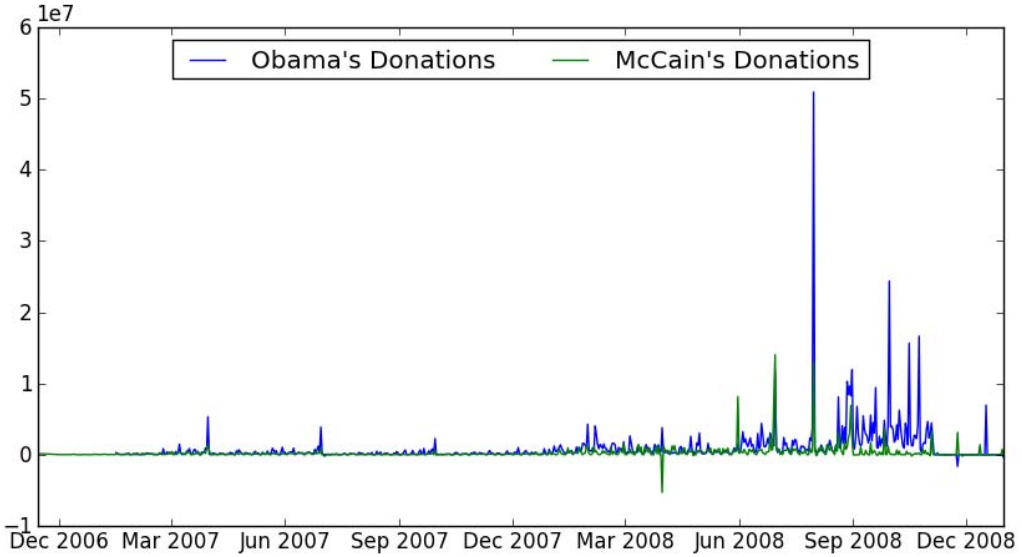
Exploratory Data Analysis

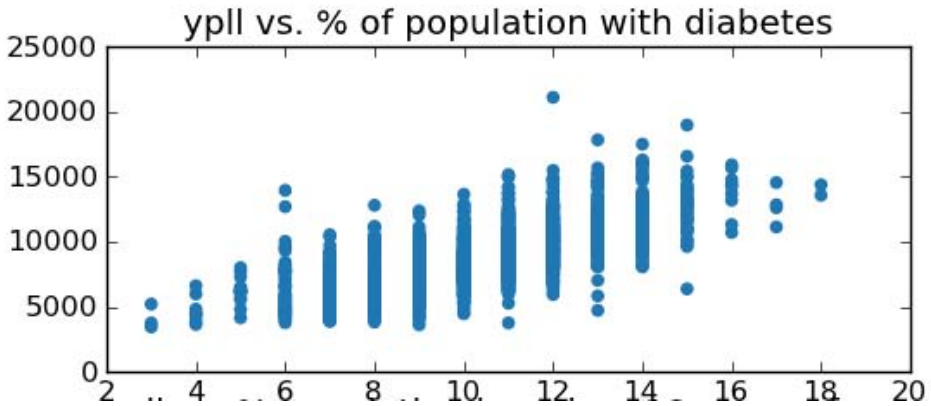
↓

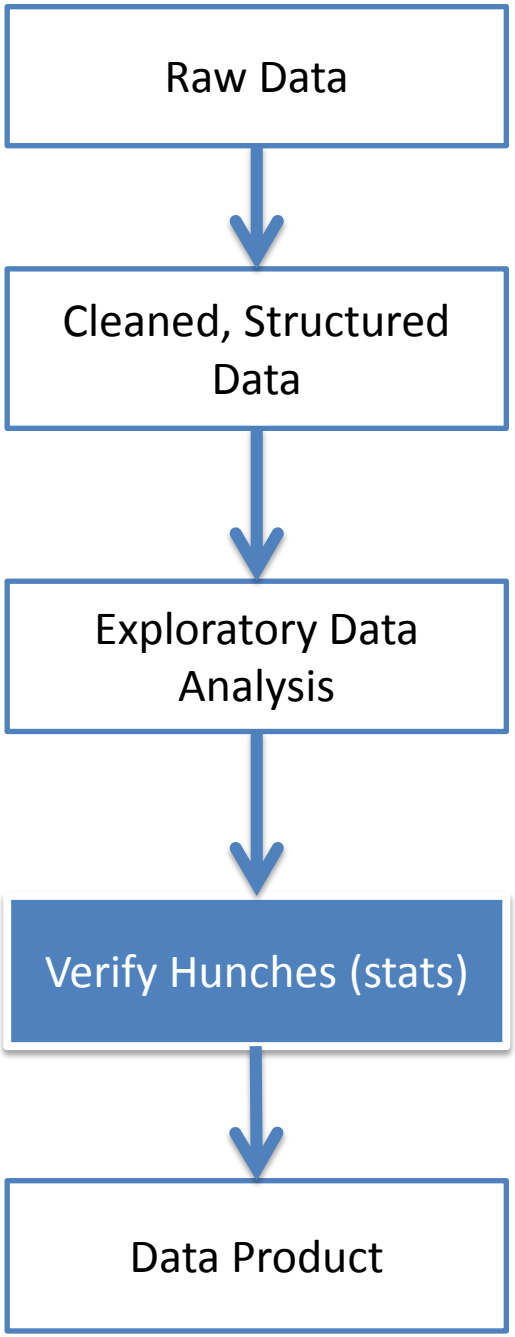
Verify Hunches (stats)

↓

Data Product

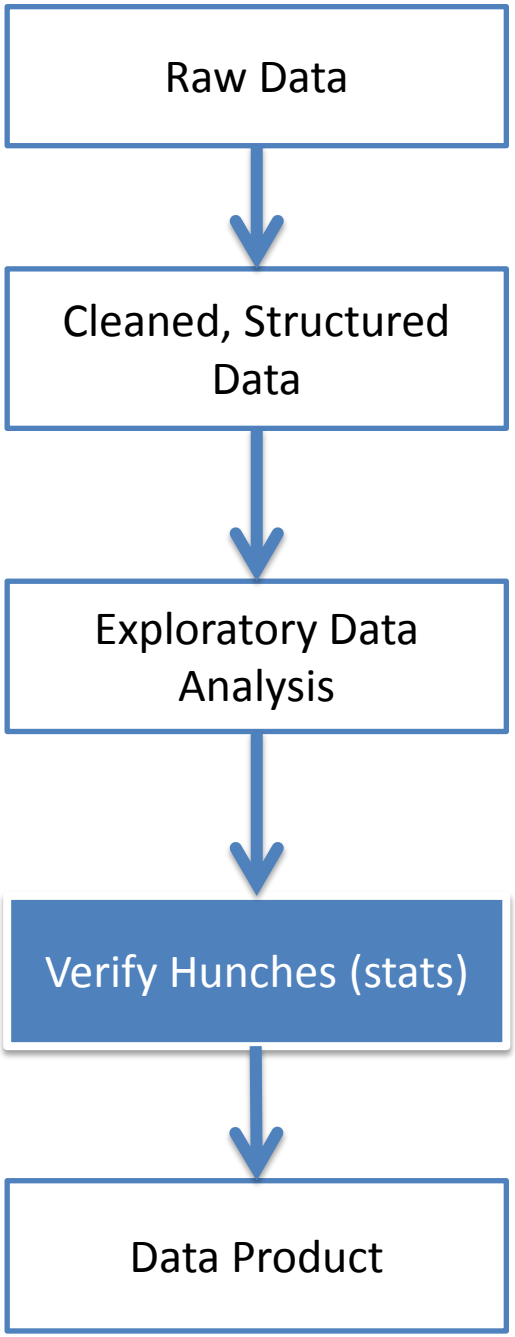






T-test

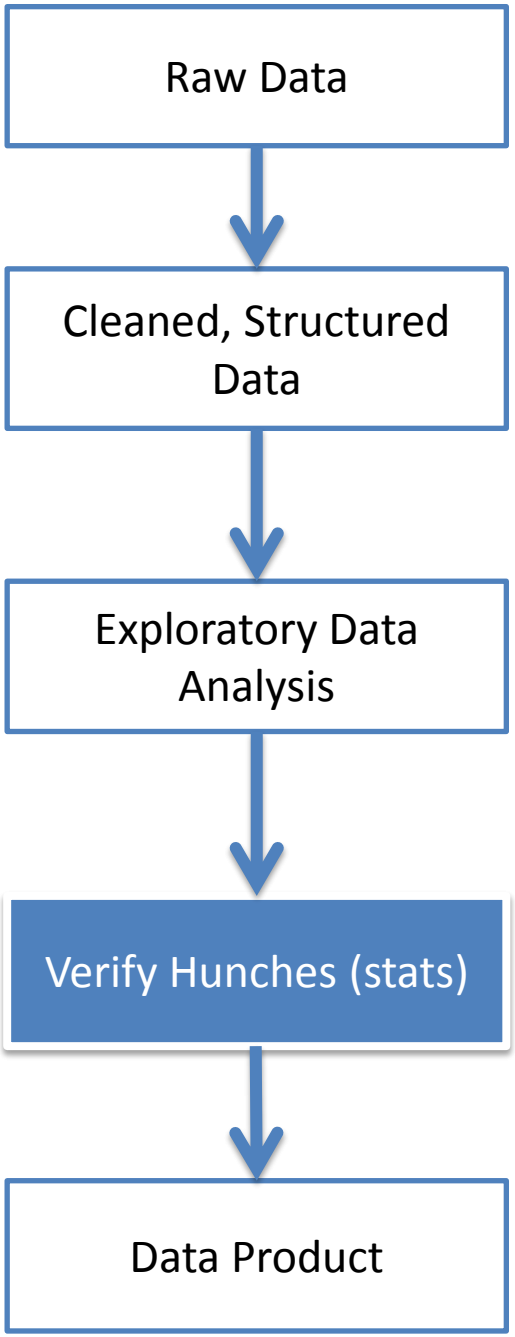
Create a model (linear regression)



T-test

Create a model (linear regression)

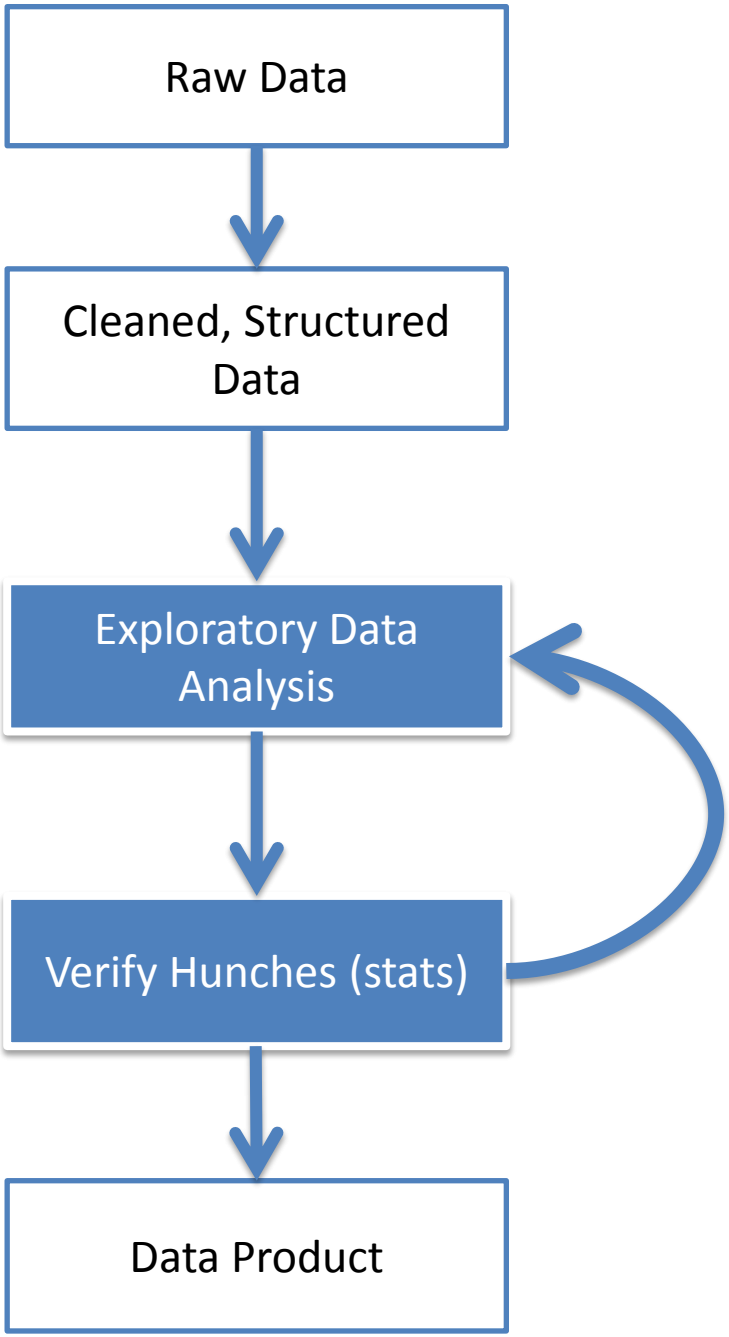
Significance



T-test

Create a model (linear regression)

~~Significance~~


Raw Data
Cleaned, Structured Data
Exploratory Data Analysis
Verify Hunches (stats)
Data Product

Raw Data

↓

Cleaned, Structured Data

↓

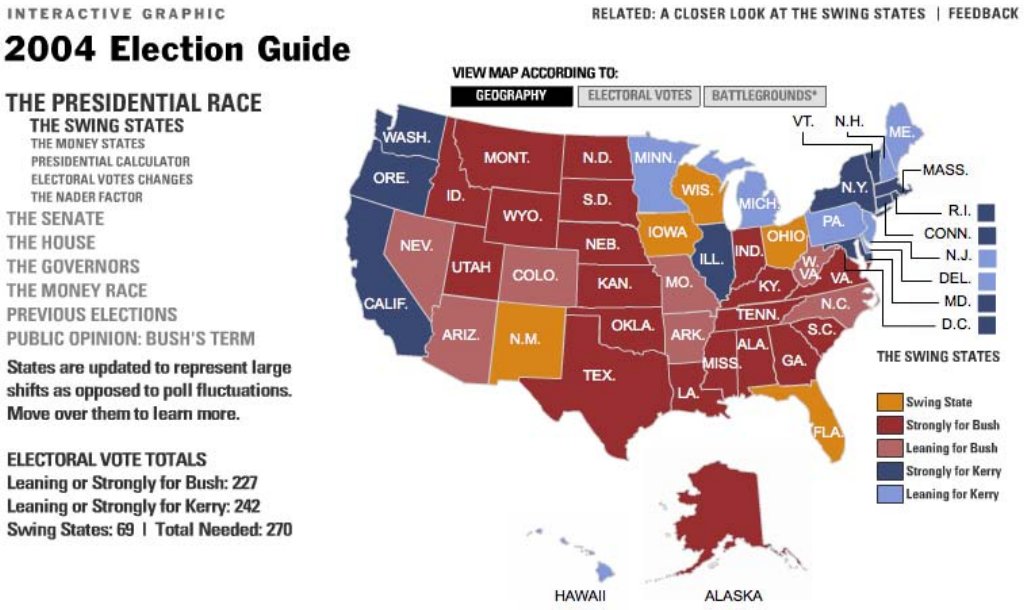
Exploratory Data Analysis

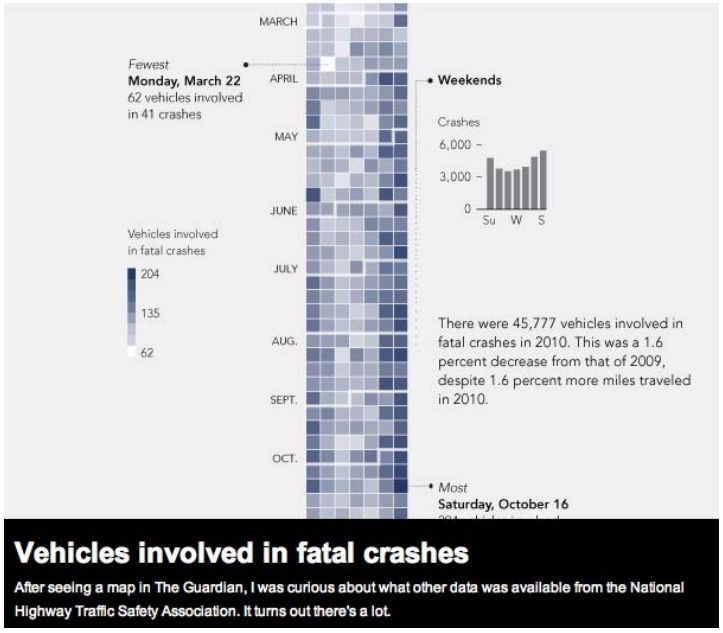
↓

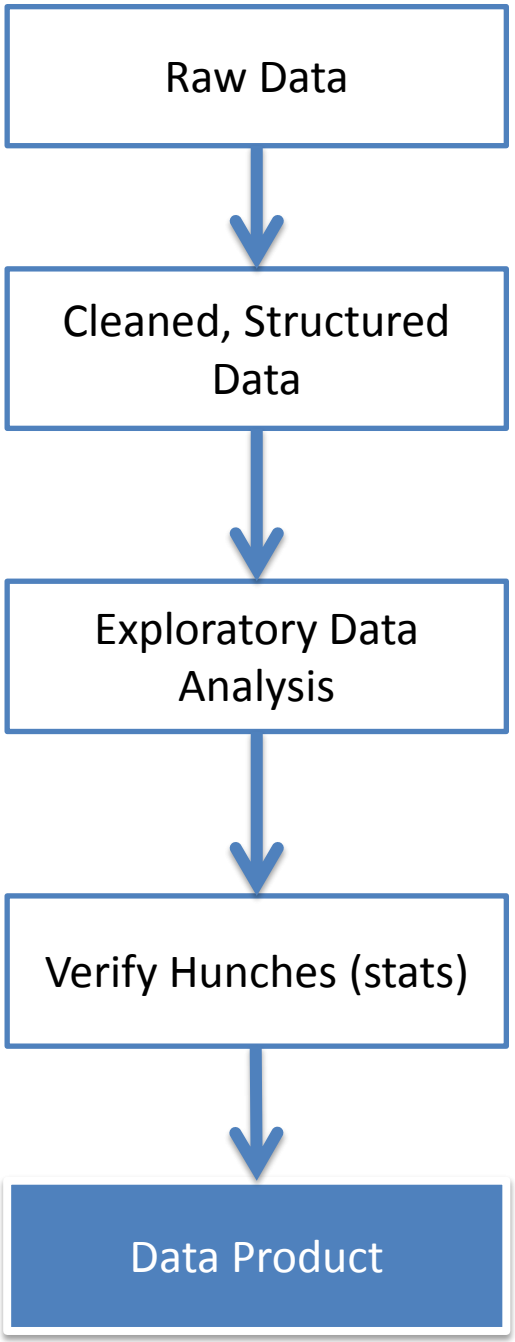
Verify Hunches (stats)

↓

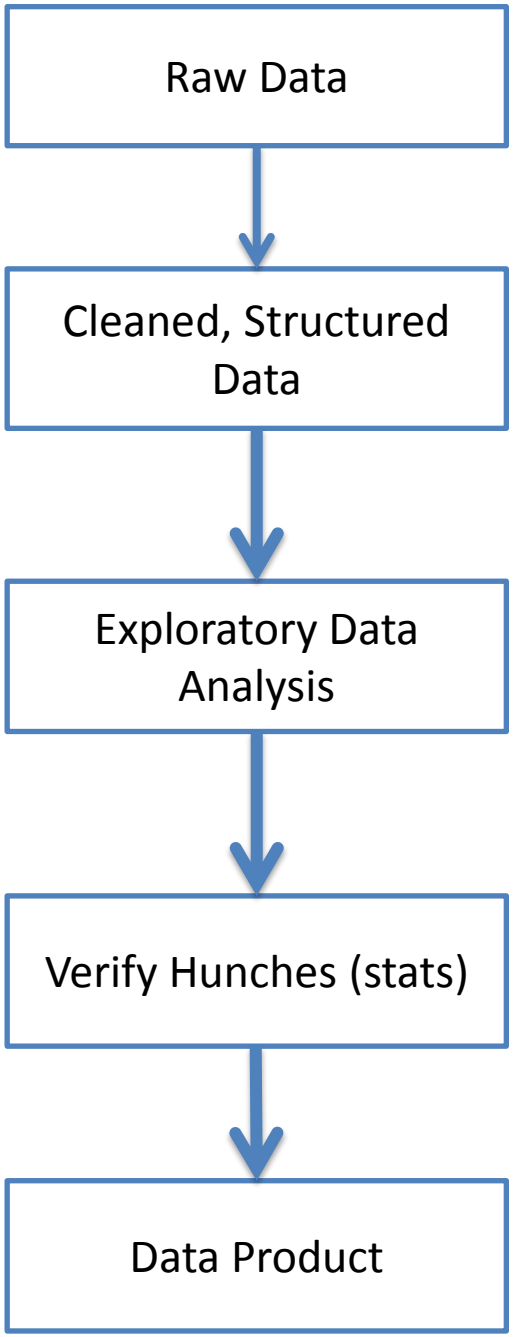
Data Product





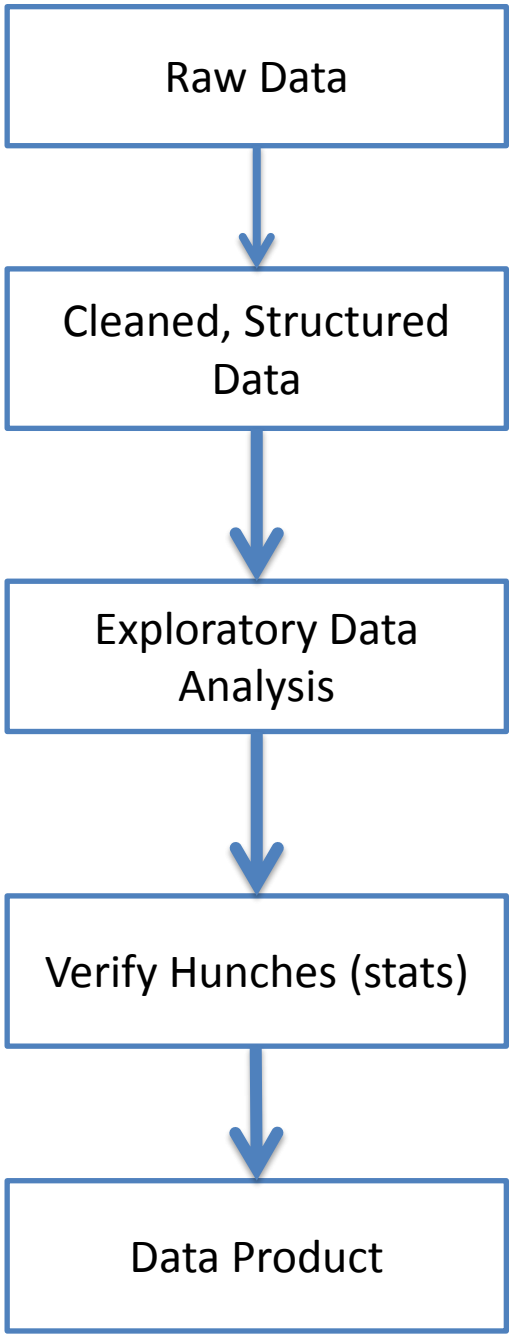


Images removed due to copyright restrictions: suggested movies on Netflix, Facebook search, LinkedIn logo.



Context

Yesterday and today, 3 companies kindly came to talk about their technologies. I personally found it awesome as well because it gives context to the stuff we've been teaching and learning.



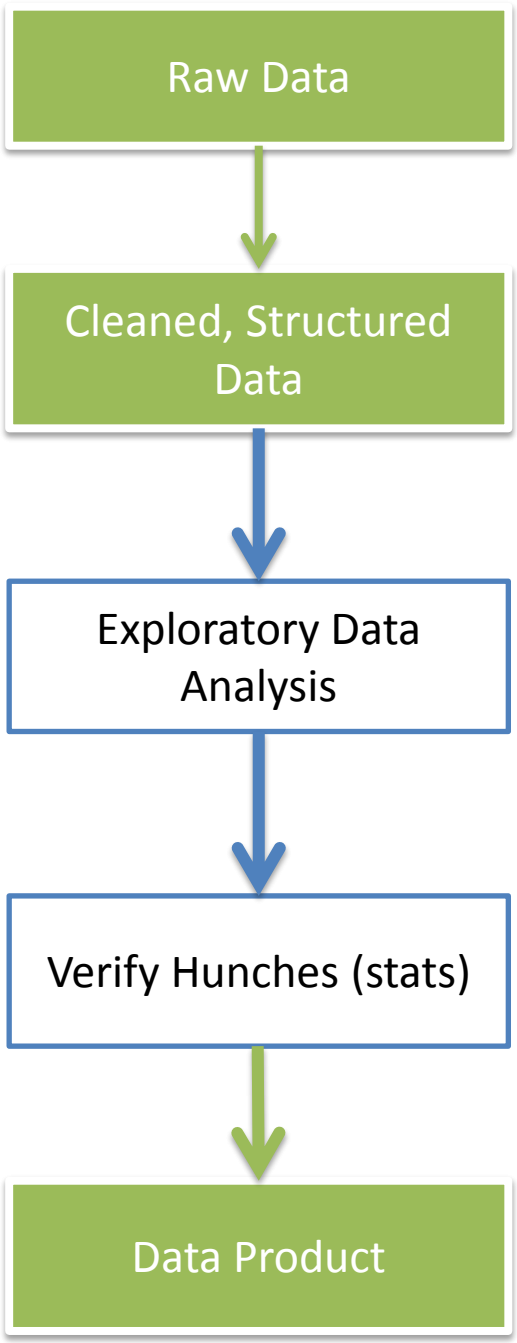
Similar

What struck me was how similar their processes are to what we've done in this class, but on a different dataset, or different scale, etc.

Pipeline

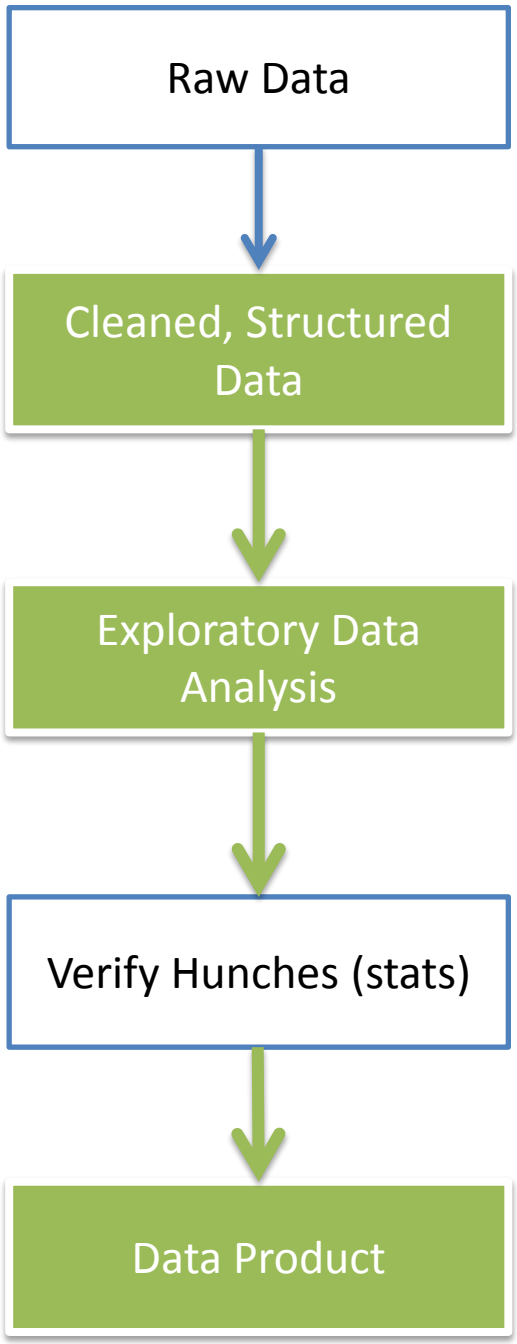
- Crazy raw data
- Cleaned, structured data
- Exploratory data analysis
- Verify Hunches
- Data Product (tm hammer@cloudera)

- Different companies fit into different subsets of the pipeline
- locu is the first segment (100% accuracy)
- visible measures is full pipeline, at huge scale
- Hadapt makes exploratory and verifying faster



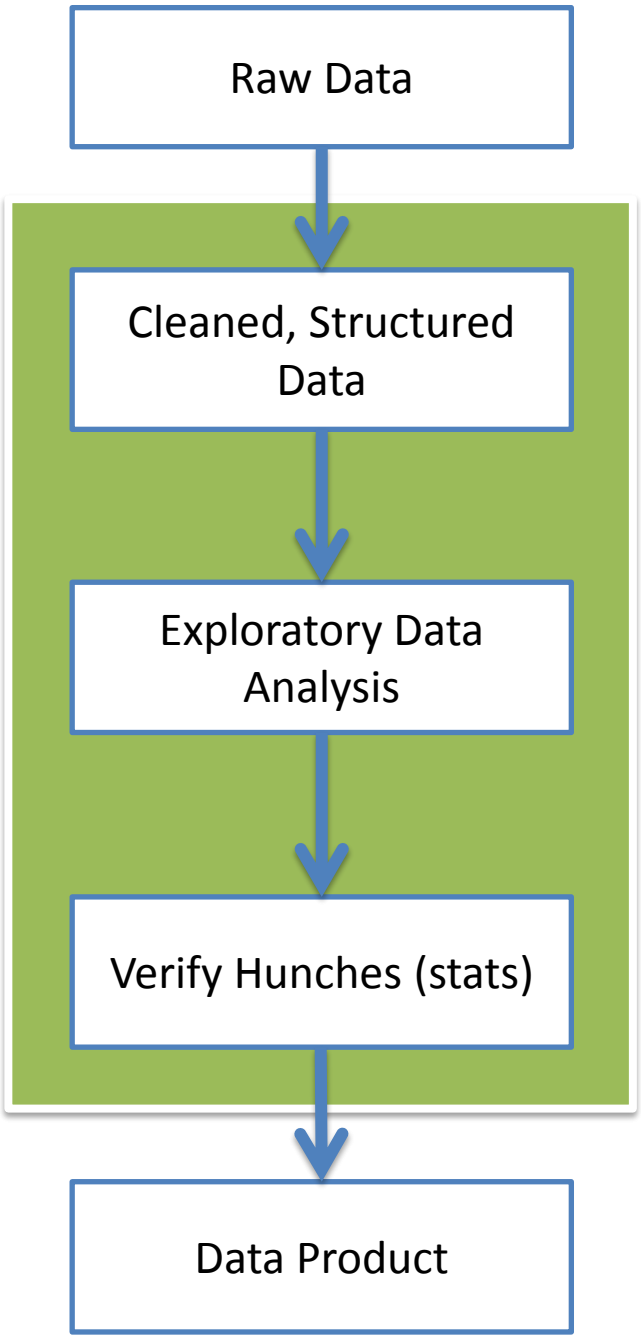
http://locu.com/

[logo removed due to copyright restrictions]



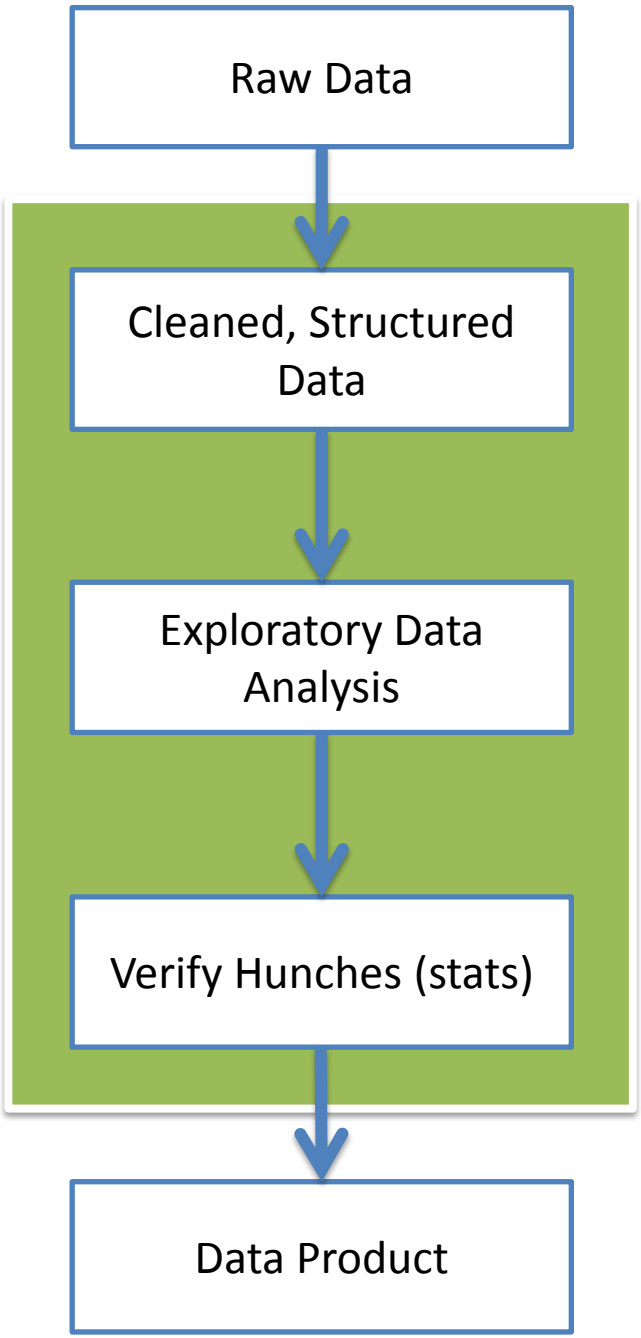
http://www.visiblemeasures.com/

Google analytics.
Takes structured apache logs (access logs) and analyzes them to see how many people are viewing a particular internet video ad.



http://www.vertica.com/

[logo removed due to copyright restrictions]



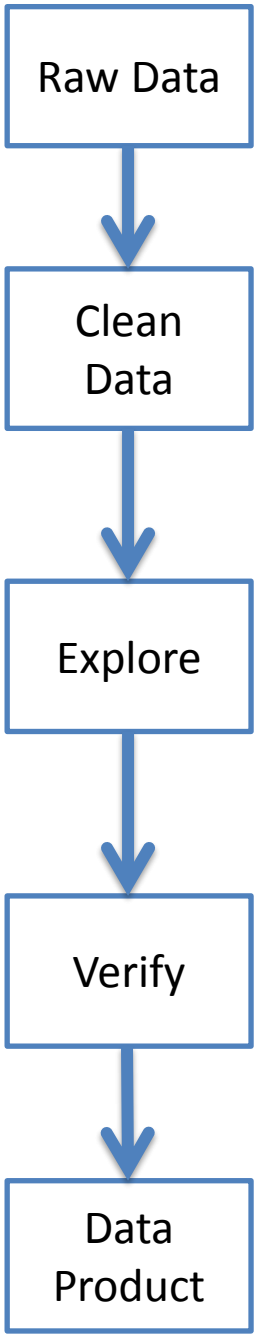
http://www.hadapt.com/

Hadapt doesn't actively perform data analysis etc. Instead, they create platforms that help other companies (like visiblemeasures) perform their data analysis faster.

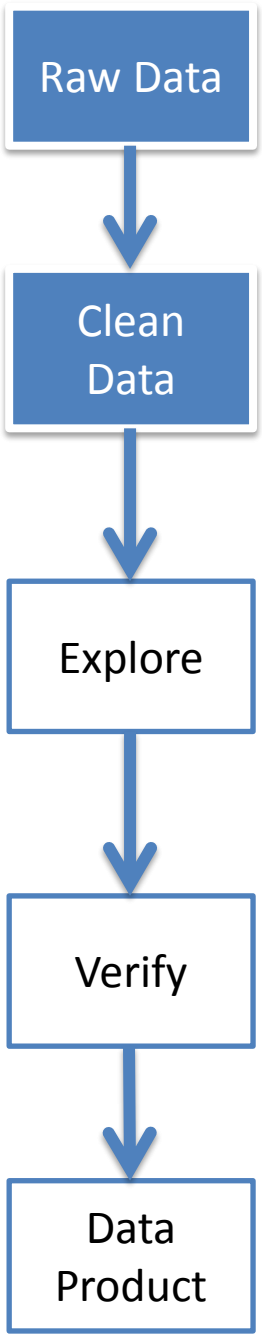
You'll find companies focused on every part of this pipeline. It's what makes companies "smarter".

- Visible Measures
- Locu

- Gave us context about what companies that are centered around data analytics are doing
- A lot of them are very similar to what we did, at a huge scale.



- Getting data
- Visualization
- Statistics
- Machine Learning
- Graph Analysis
- Text Analysis
- Databases
- “Big Data”



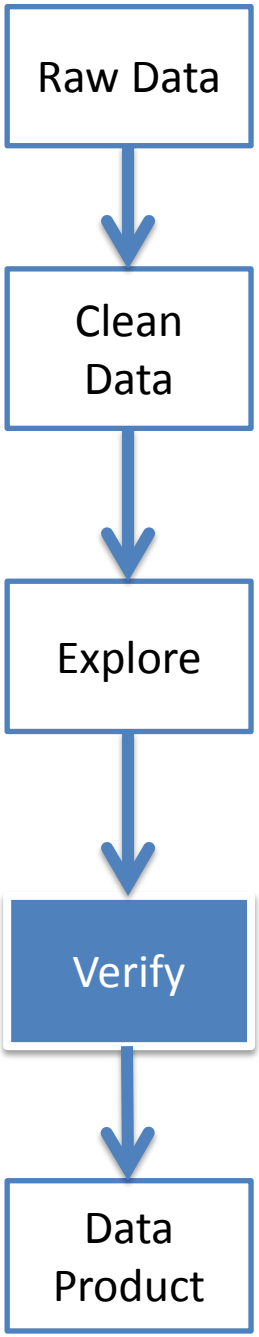
Getting Data

- Surveys
- Web Crawling/Scraping
  - https://scraperwiki.com
  - http://nutch.apache.org
- Sensors

Visualizations

- Interactive Visualizations
  - HTML5/CSS/JavaScript
- Tools
  - processingjs, d3, prefuse
- Blogs
  - http://flowingdata.com
  - http://infosthetics.com

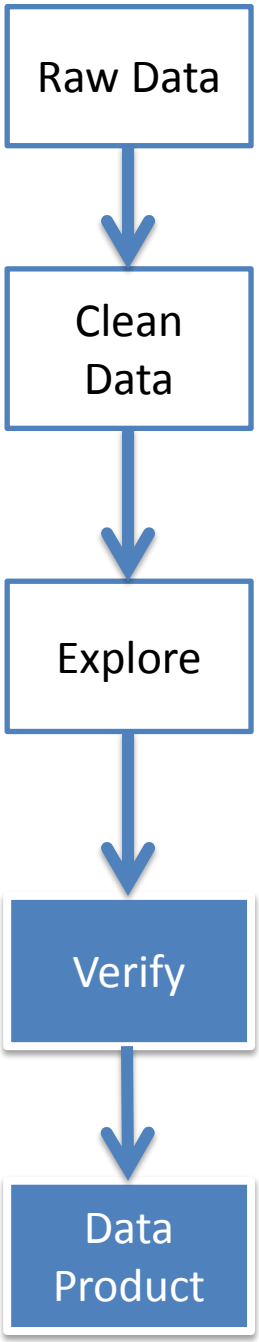
- Harvard http://cs171.org
- MIT 6.831



Statistics

- Are they different?
  - T-Tests, ANOVA
- Bayesian Statistics
- Correlation
- Regressions
  - Linear
  - Non-Linear

- 16.470j
- http://statistics.mit.edu



Machine Learning

- Classification
- Clustering

- http://www.ml-class.org
- MIT 6.867
- Python scikit-learn (sklearn)

Graph Analysis

- Examples:
  - web pages, friend graph, twitter
- Metrics
  - Centrality
  - Cohesion
  - “Importance” (page rank)

- Social Network Analysis
- Web data mining MIT Course
  - Sep Kamvar Fall 2012
- http://www.stats.ox.ac.uk/~snijders/sna_course.htm

Text Analysis

- Natural Language Processing
  - Parsing sentences
  - Extracting the grammar/structure
- Similarity measures
  - Cosine Similarity
  - Jaccard
- Identifying Entities
  - Opencalais
- MIT 6.864/6.863J

Databases

- SQL Implements a lot of what we did
  - Filtering
  - Joining
  - Grouping
  - Summarizing
- Specialized system to do this
  - SQL databases, Hive, Pig

- MIT 6.830
- http://db-class.org

"Big Data"

- How to process on 1000+ machines?
- Problems
  - Managing
  - Machines fail all the time
  - Network problems
  - Data out-of-sync (consistency)

- Distributed Systems
- MIT 6.824 (6.830 a bit)

Berkeley Also Has a Class!

http://datascienc.es

Thank You!

git pull

Resource: How to Process, Analyze and Visualize Data

Adam Marcus and Eugene Wu