OK Google, Tell Me About Myself

Lisa Chang

About Me

- ~~Process Contact Engineer~~
- ~~Process Development Engineer~~
- ~~Applications Engineer~~
- ~~Scientist~~
- Software Engineer
- Data Scientist

2018 Data Headlines

2018 reform of EU data protection rules

Stronger rules on data protection mean people have more control over their personal data and businesses benefit from a level playing field.

PAGE CONTENTS

Rules for business and organisations

Application of the GDPR obligations, individuals' requests, enforcement

Rights for citizens

Protection of your personal data, your rights and redress

About the regulation and data protection

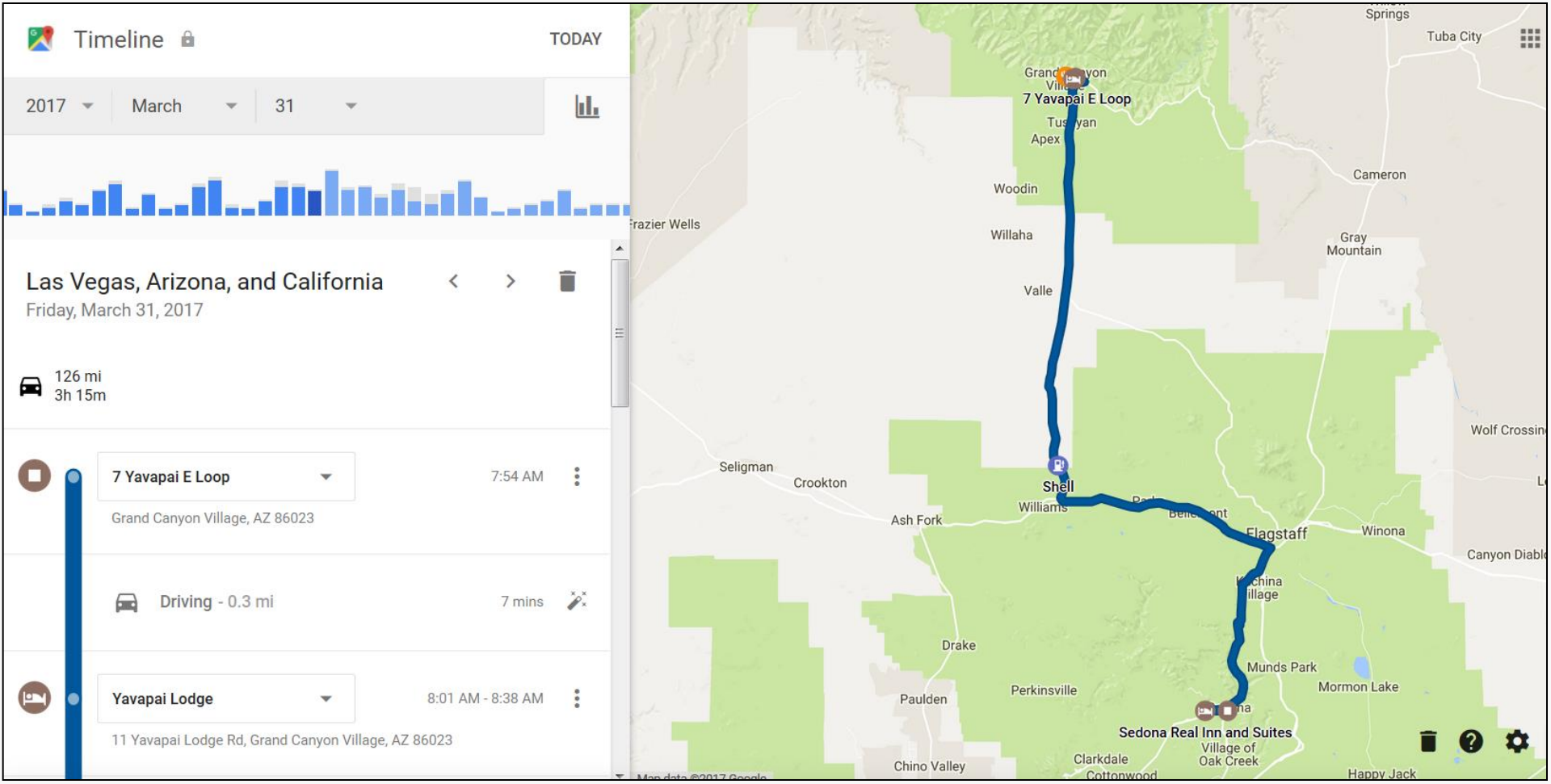
- What does the General Data Protection Regulation (GDPR) govern?
- What is personal data?
- What constitutes data processing?

TECHNOLOGY

3 Social Media Data Lessons in the Wake of Cambridge Analytica

More transparency and higher awareness of how data is used can only be good.

Google Timelines



Your July in review

Your timeline in Google Maps helps you curate the places you've been. Look back on the past month and reminisce about recent trips and past places.

EXPLORE YOUR TIMELINE

8 cities visited this month



28 places visited this month (5 new)



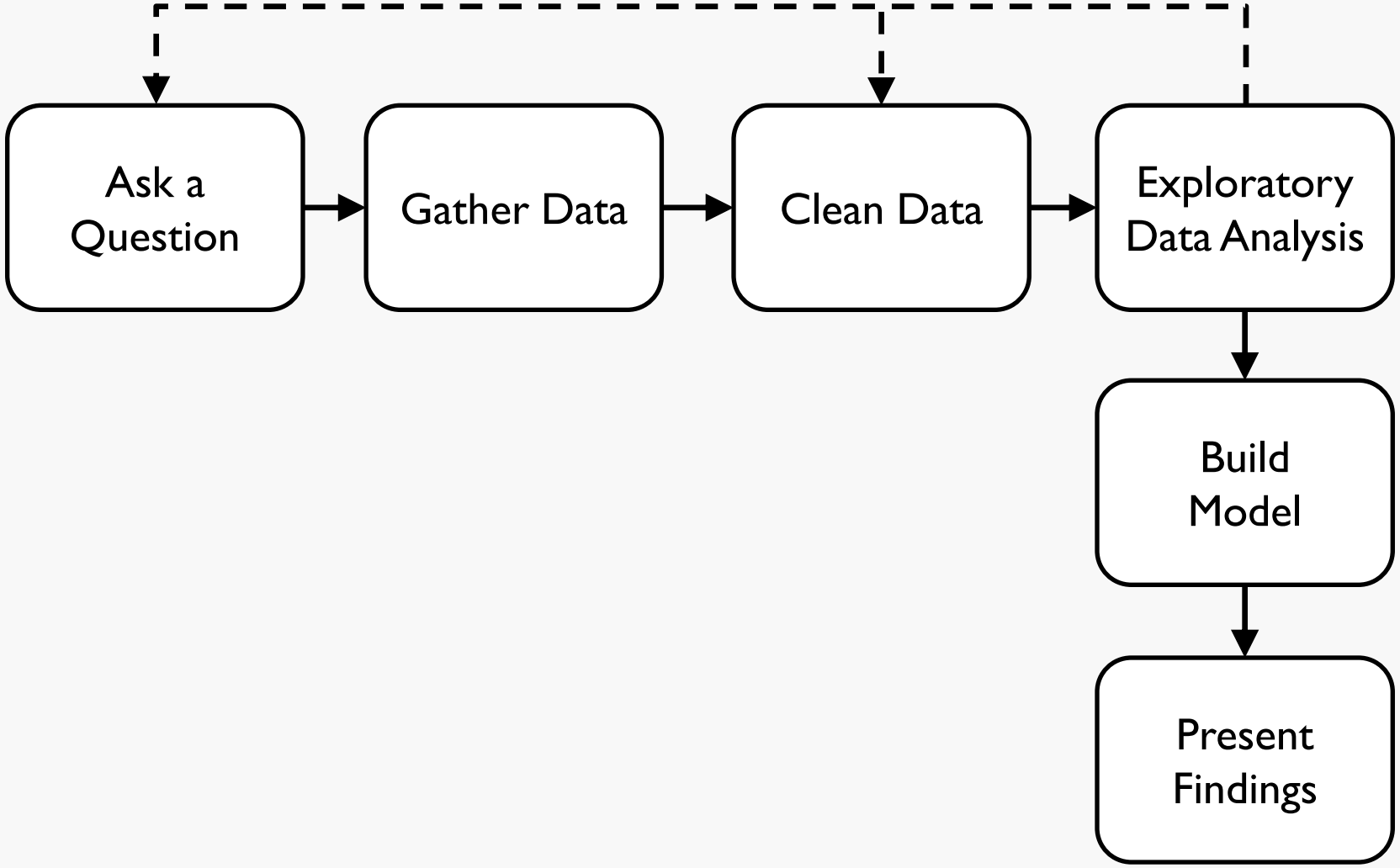
Your activity in timeline

6 mi (11 km)
walked this month

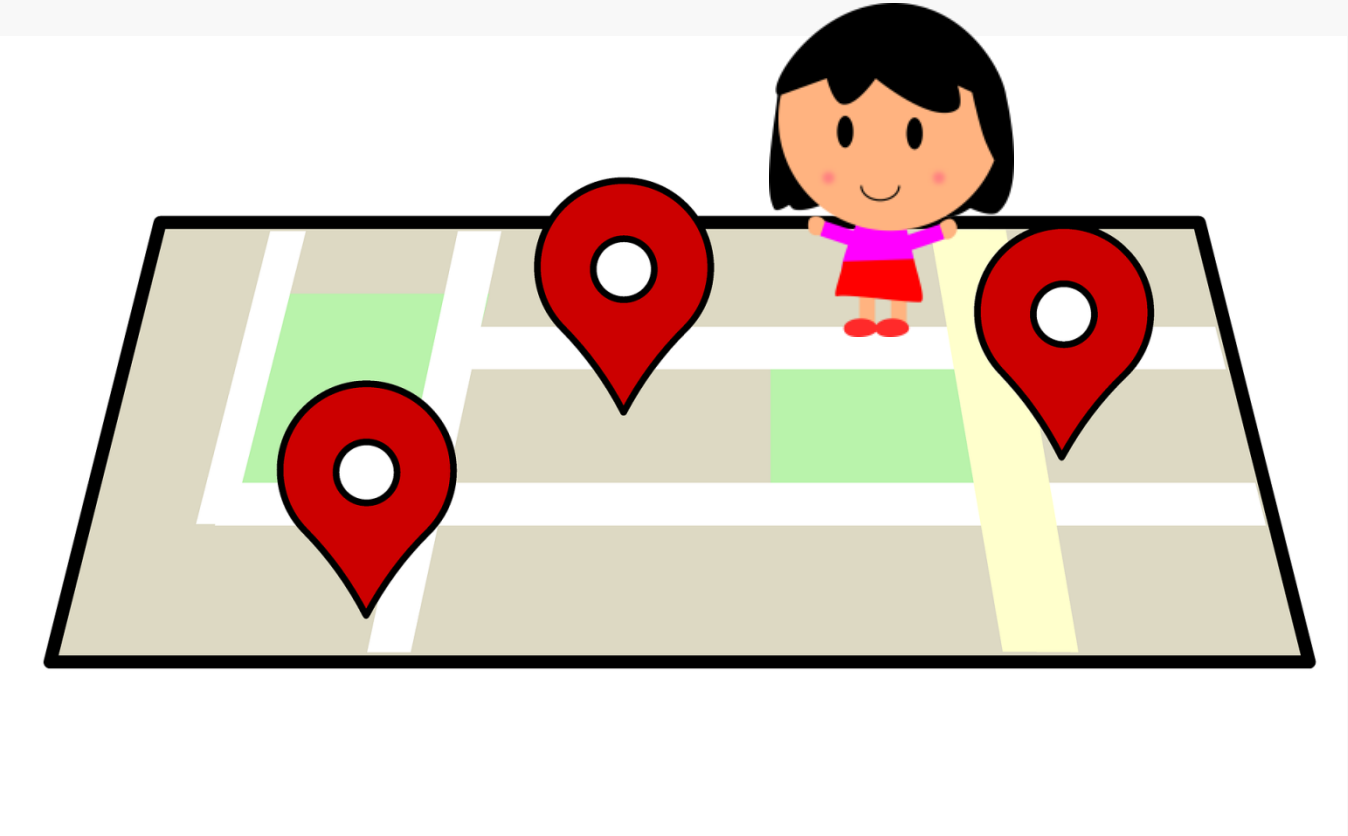
22 mi (37 km)
run this month

28 hours spent in a vehicle this month

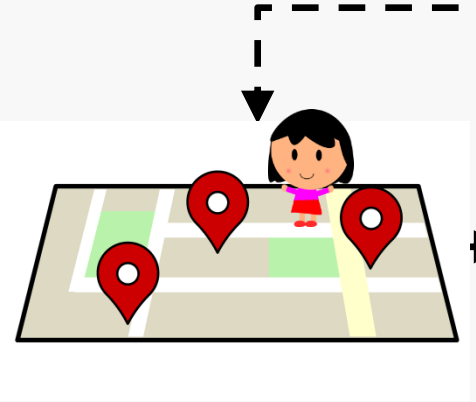
Data Science Process



Can I create a model of my life from stored location data?



Data Science Process



Gather Data

Clean Data

Exploratory Data Analysis

Build Model

Present Findings

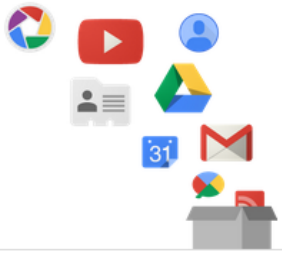
https://takeout.google.com

Download your data

Your account, your data.
Export a copy.

Create an archive with your data from Google products.

Manage archives



Location History	JSON format	✓
Mail	All mail	✓
Maps (your places)		✓
My Maps		✓
Searches		✓
Tasks		✓
Voice		✓
YouTube	All data types OPML (RSS) format	✓

KML format example

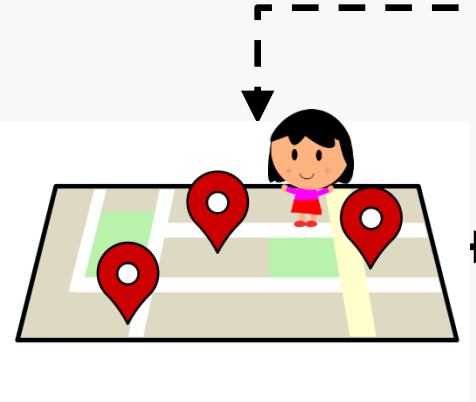
```
<when>2017-03-30T22:16:05Z</when>
<gx:coord>-112.1206089 36.0538447 2110</gx:coord>

<when>2017-03-30T22:15:32Z</when>
<gx:coord>-112.1206895 36.0541252 2108</gx:coord>

<when>2017-03-30T22:14:41Z</when>
<gx:coord>-112.1161455 36.0566548 2117</gx:coord>

<when>2017-03-30T22:13:41Z</when>
<gx:coord>-112.1110006 36.0585582 2123</gx:coord>
```

JSON format example

Not always available


```
"timestampMs" : "1490998907806",
"latitudeE7" : 348600316,
"longitudeE7" : -1118161027,
"accuracy" : 21,
"activity" : [ {
  "timestampMs" : "1490998831576",
  "activity" : [ {
    "type" : "STILL",
    "confidence" : 75
  }, {
    "type" : "ON_FOOT",
    "confidence" : 10
  }, {
    "type" : "IN_VEHICLE",
    "confidence" : 5
  }, {
    "type" : "ON_BICYCLE",
    "confidence" : 5
  }, {
    "type" : "UNKNOWN",
    "confidence" : 5
  }, {
    "type" : "WALKING",
```

Data Science Process



Gather Data

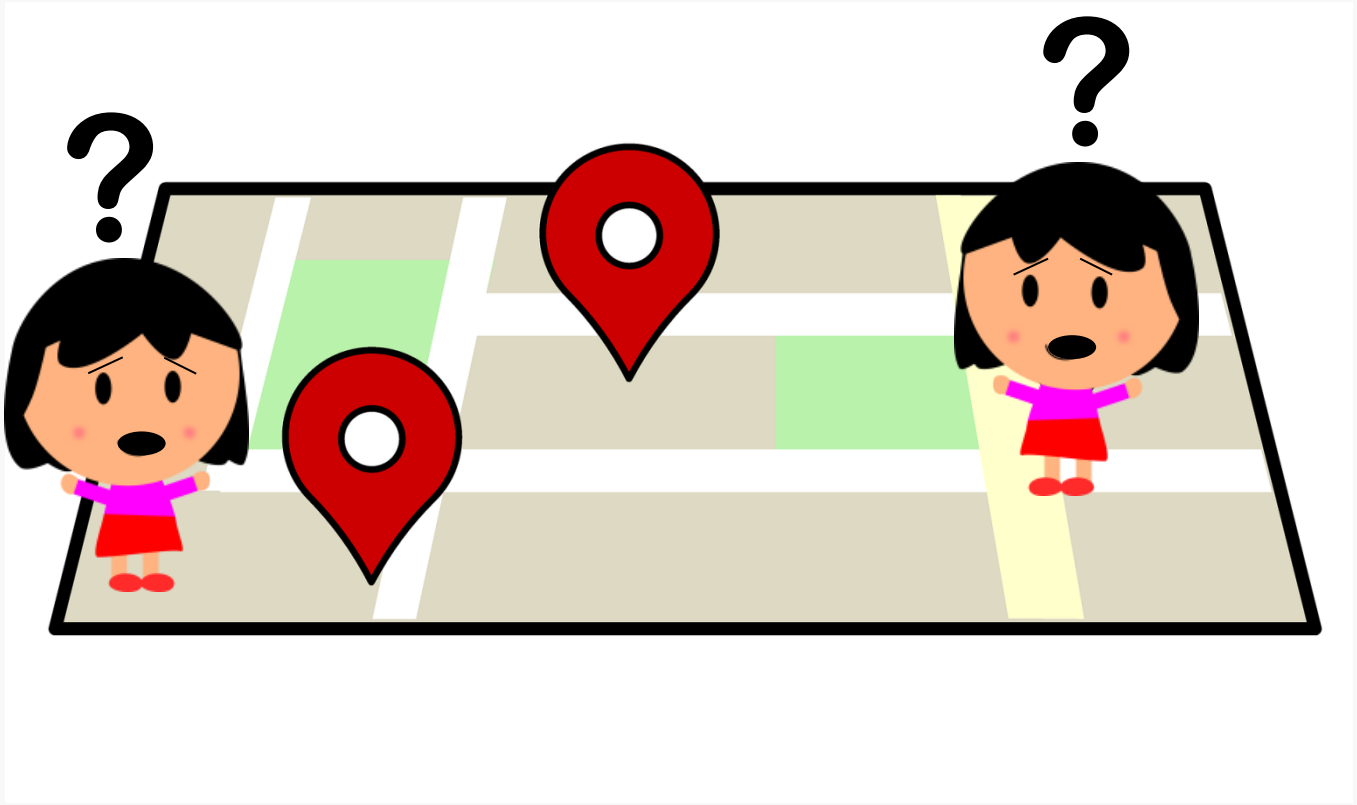
Clean Data

Exploratory Data Analysis

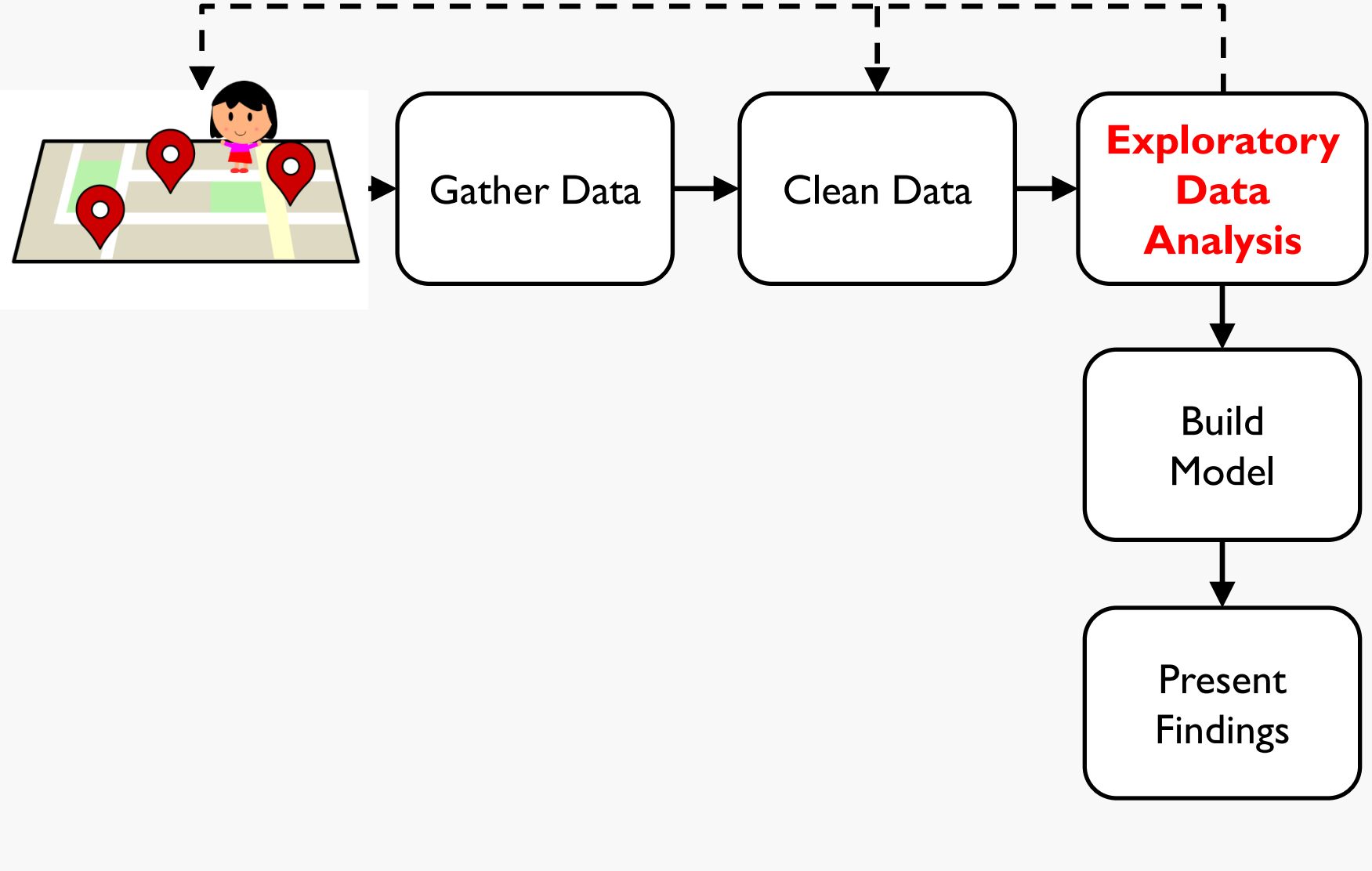
Build Model

Present Findings

Traveling at the speed of light



Data Science Process

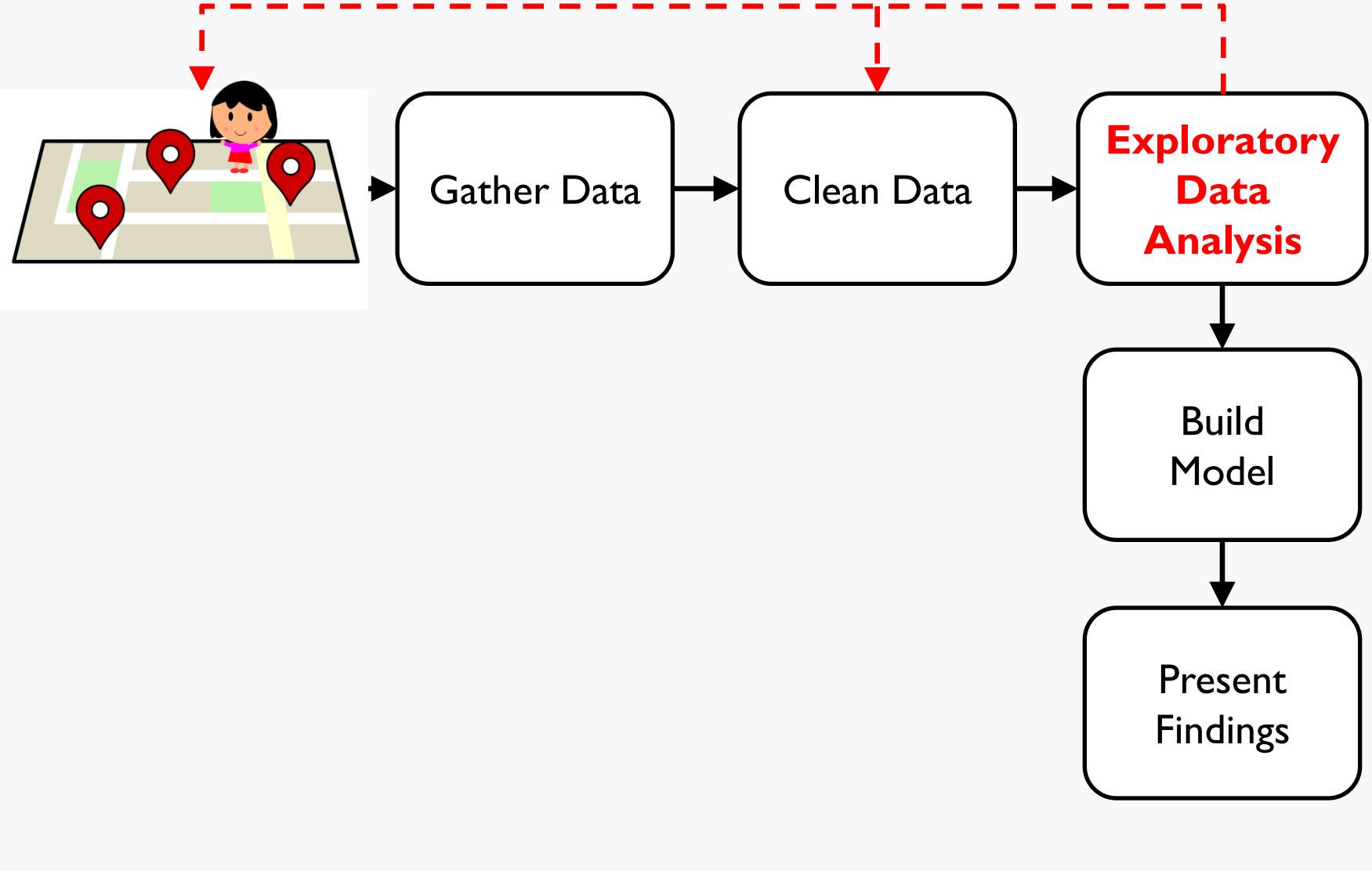


What is EDA?

- Define characteristics
  - Trends
  - Biases
  - Variability
  - Breadth

- Test Assumptions
- Visualize

Data Science Process



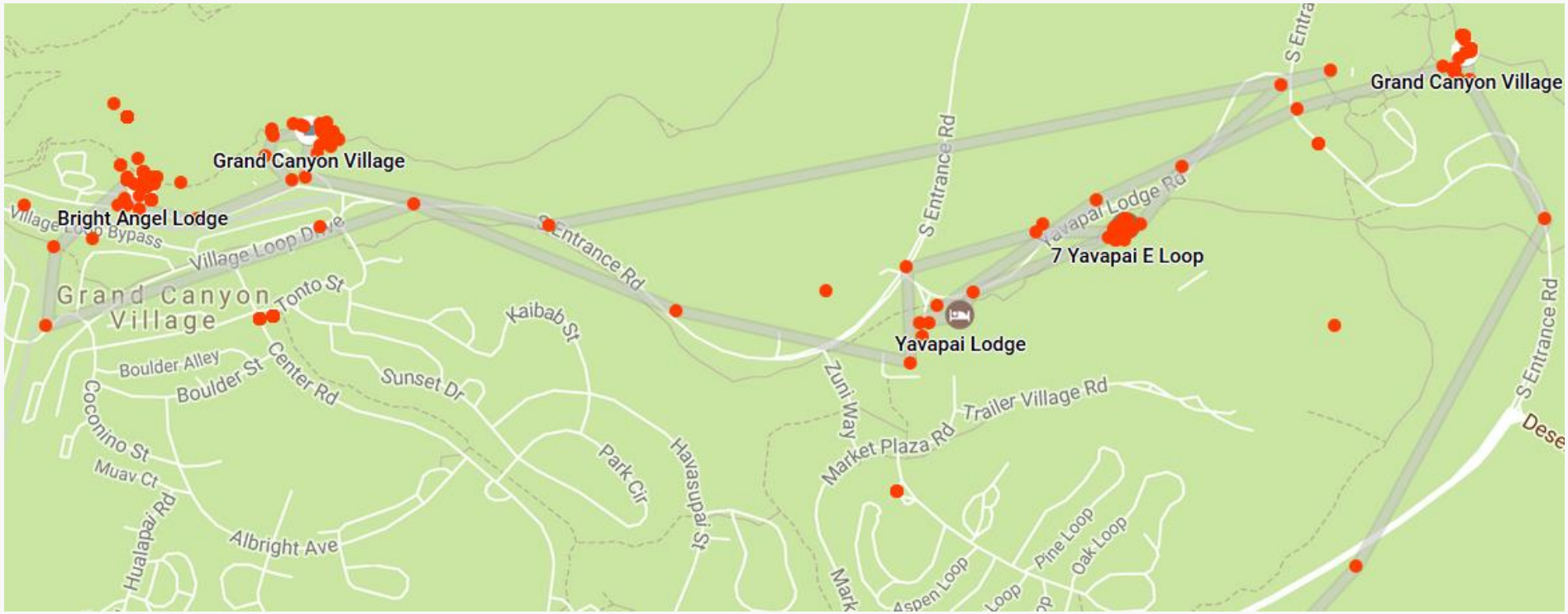
What's in the data?

```
<when>2017-03-30T22:16:05Z</when>
<gx:coord>-112.1206089 36.0538447 2110</gx:coord>

<when>2017-03-30T22:15:32Z</when>
<gx:coord>-112.1206895 36.0541252 2108</gx:coord>

<when>2017-03-30T22:14:41Z</when>
<gx:coord>-112.1161455 36.0566548 2117</gx:coord>

<when>2017-03-30T22:13:41Z</when>
<gx:coord>-112.1110006 36.0585582 2123</gx:coord>
```

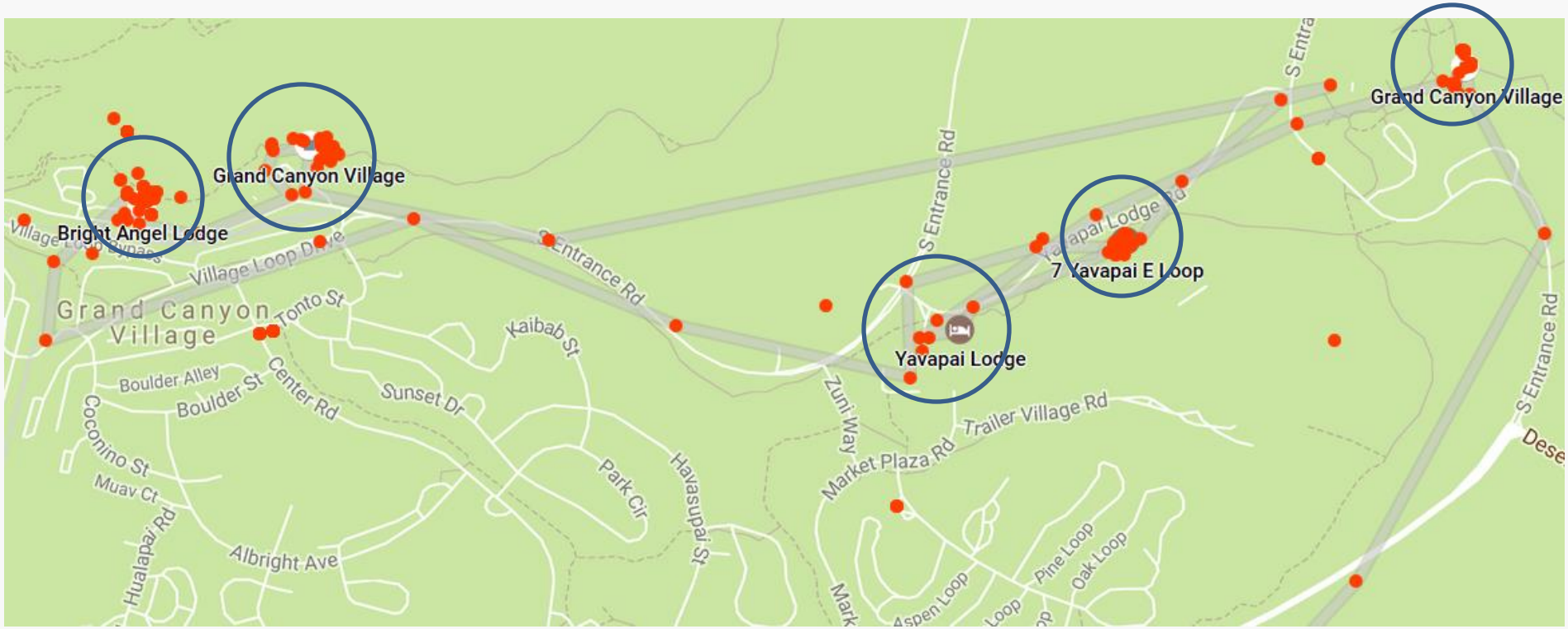
Location

```
<when>2017-03-30T22:16:05Z</when>
<gx:coord>-112.1206089 36.0538447 2110</gx:coord>

<when>2017-03-30T22:15:32Z</when>
<gx:coord>-112.1206895 36.0541252 2108</gx:coord>

<when>2017-03-30T22:14:41Z</when>
<gx:coord>-112.1161455 36.0566548 2117</gx:coord>

<when>2017-03-30T22:13:41Z</when>
<gx:coord>-112.1110006 36.0585582 2123</gx:coord>
```

Grand Canyon

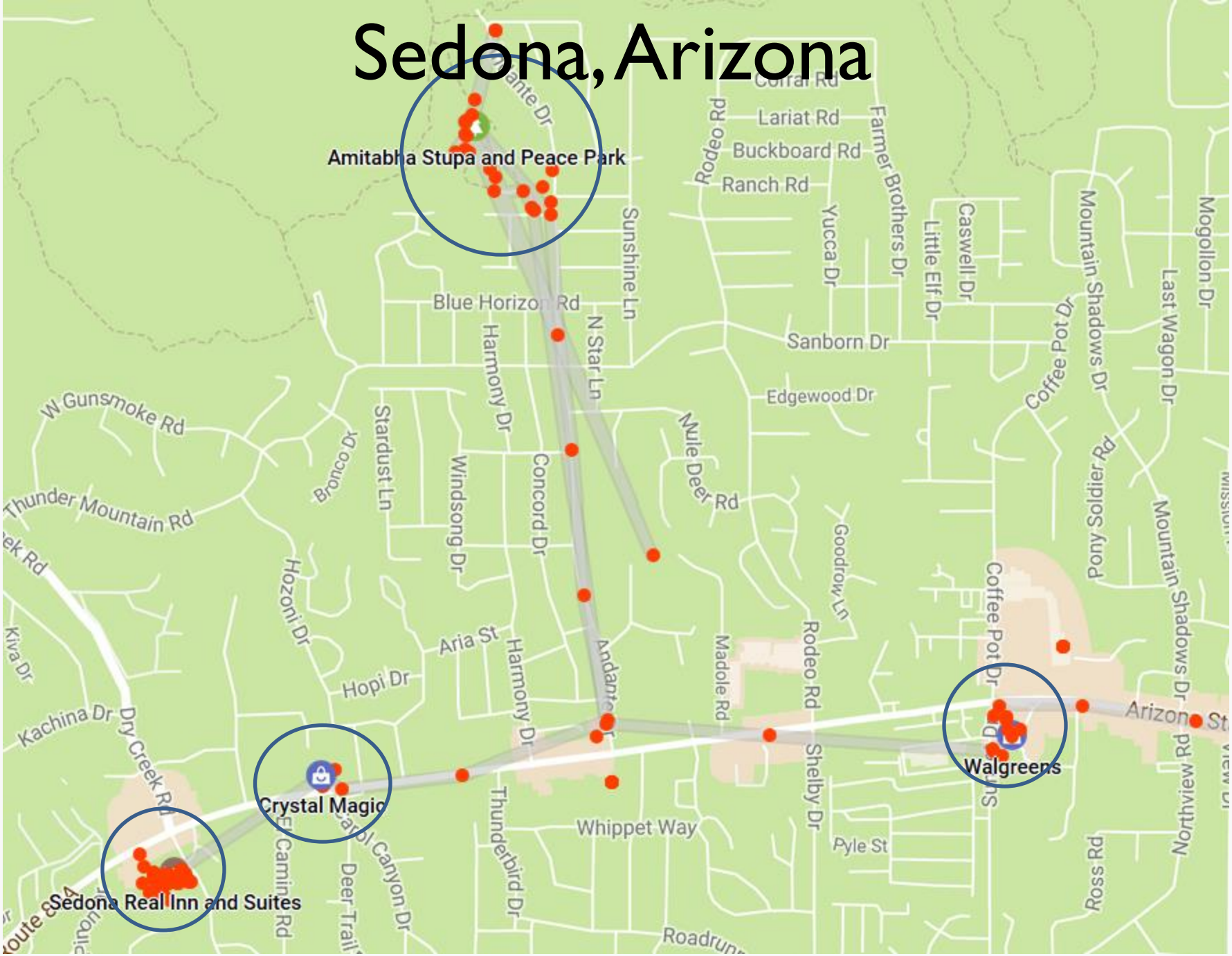


Grand Canyon

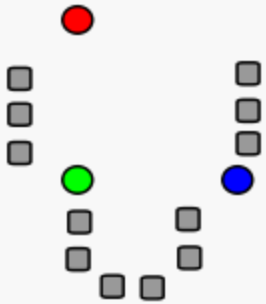


Sedona, Arizona

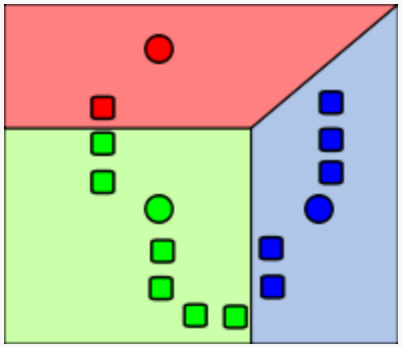
Sedona, Arizona



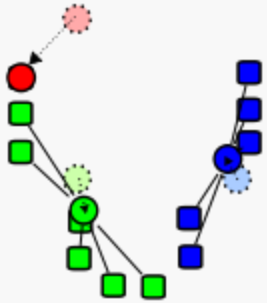
K-Means



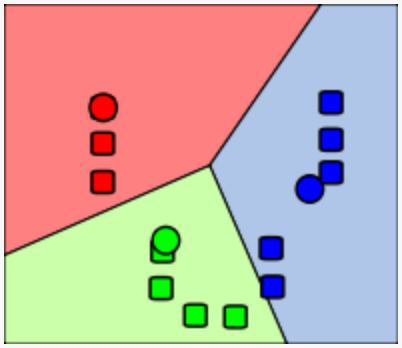
Randomly pick K = 3 points (initial centroids)



Assign each point to its closest centroid



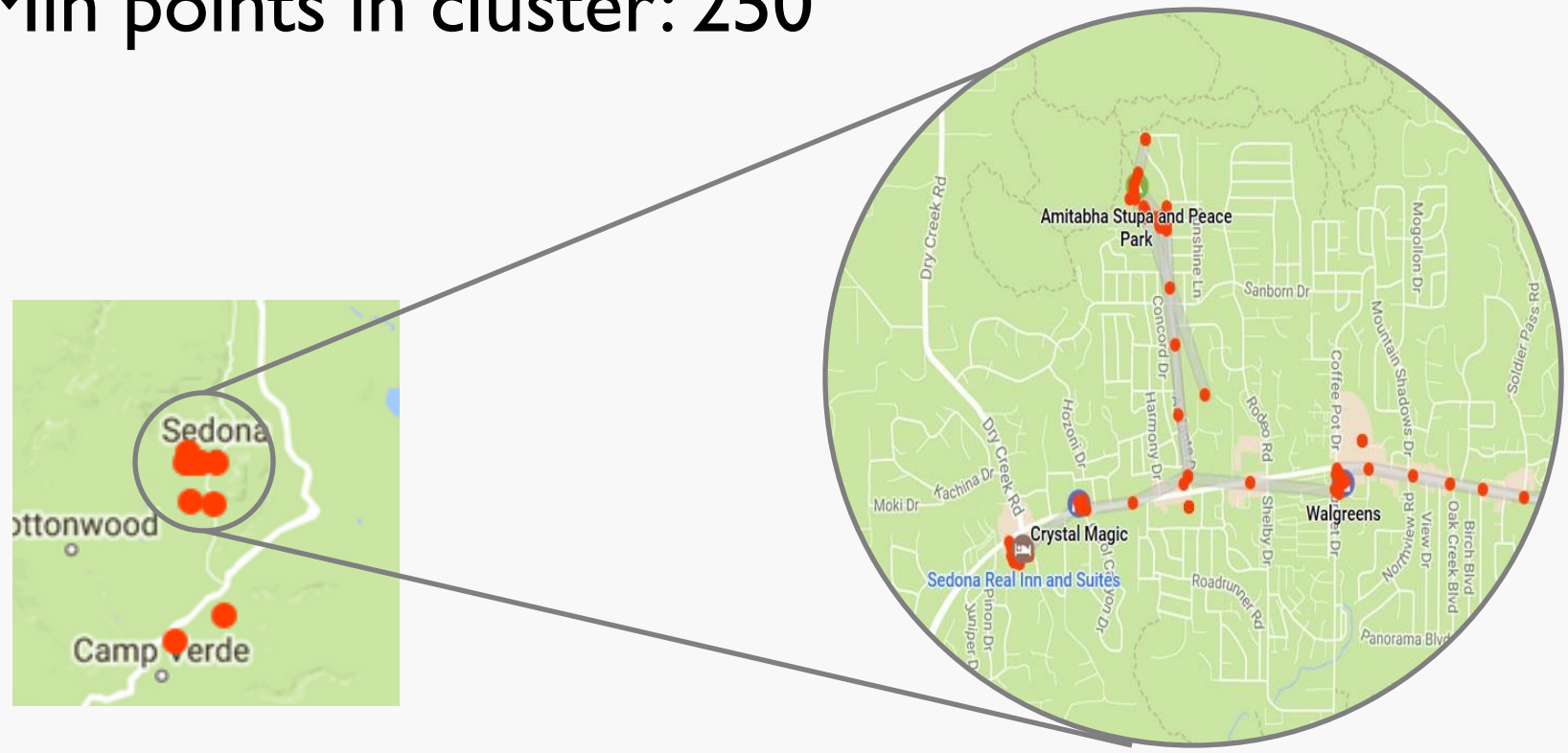
Using points in clusters, calculate new centroids

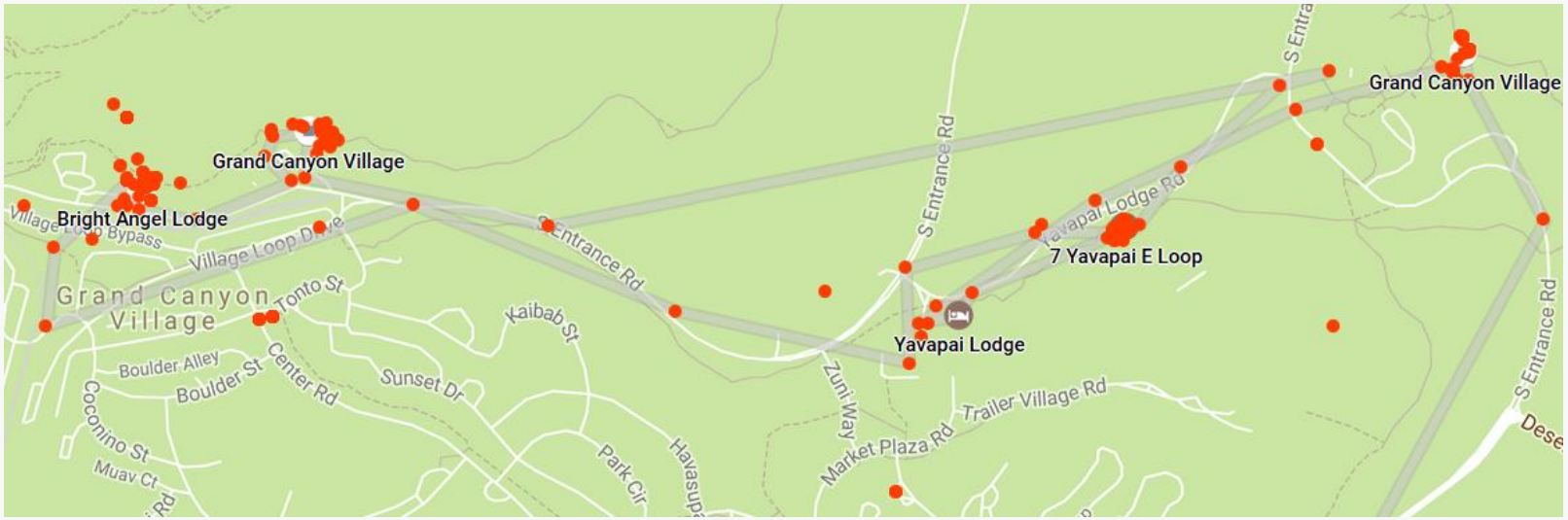


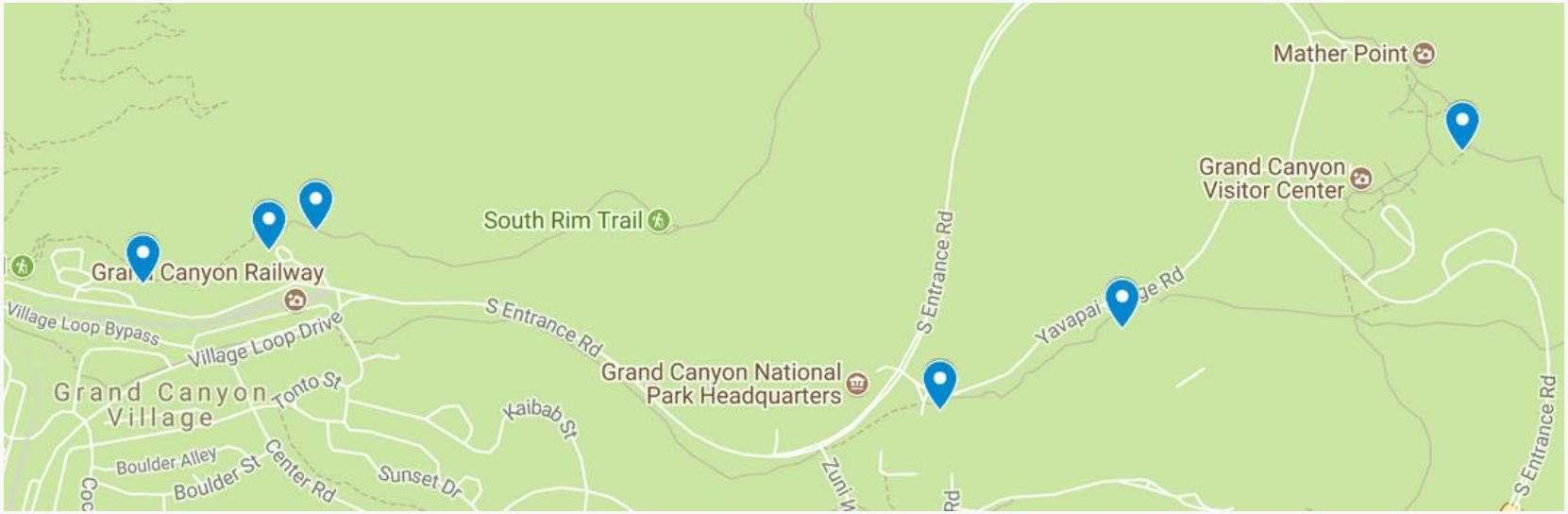
Assign each point to its closest centroid

Recursive K-Means

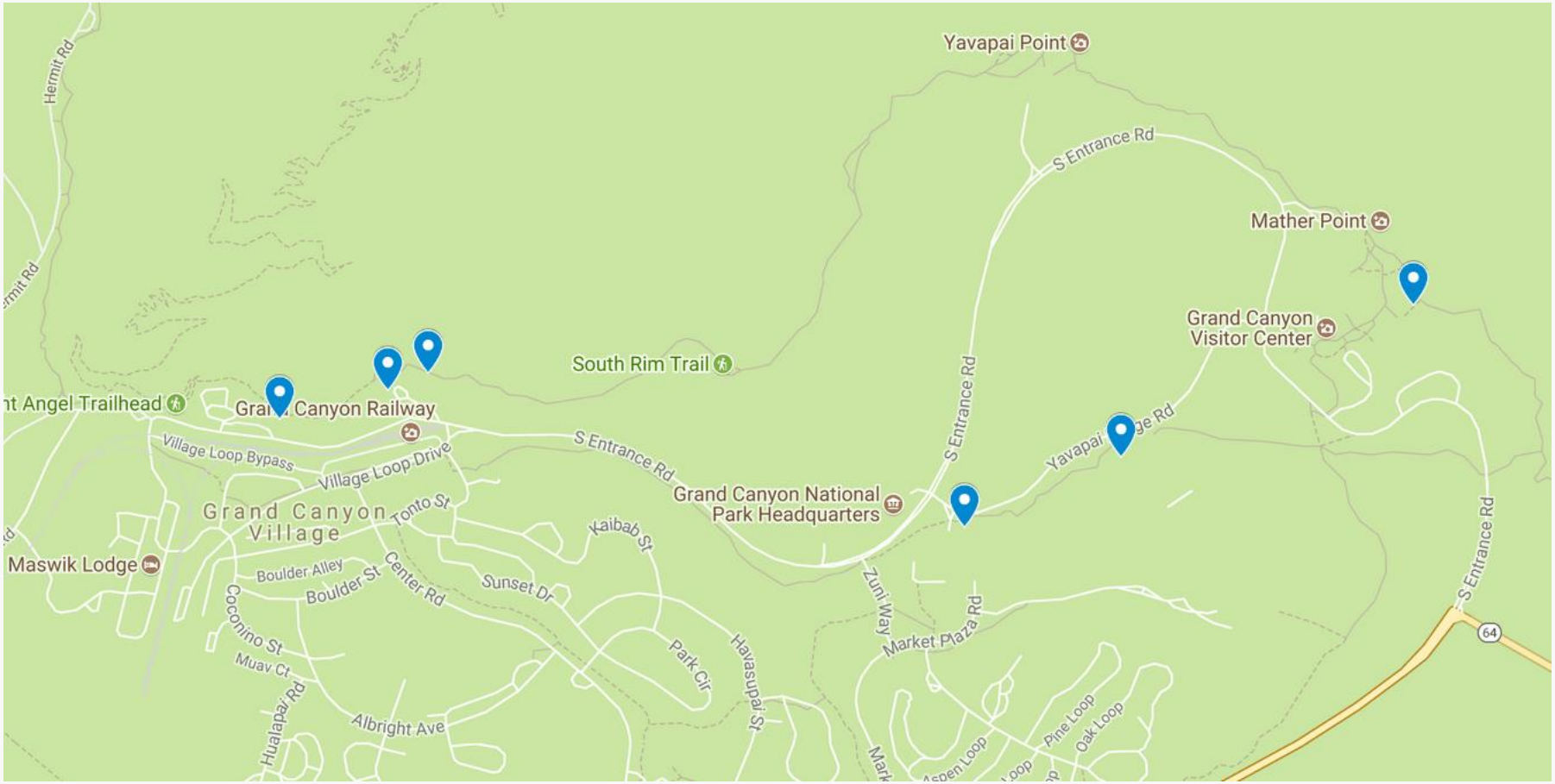
- Max radius: 0.1 miles
- Min points in cluster: 250




Grand Canyon Village
Bright Angel Lodge
Grand Canyon Village
Village Loop Bypass
Village Loop Drive
S Entrance Rd
Tonto St
Grand Canyon Village
Kaibab St
Boulder Alley
Boulder St
Center Rd
Sunset Dr
Coconino St
Muav Ct
Park Cir
Havasupai
Zuni Way
Market Plaza Rd
Trailer Village Rd
Yavapai Lodge Rd
7 Yavapai E Loop
Yavapai Lodge
S Entrance Rd
Grand Canyon Village
S Entrance Rd


Mather Point
Grand Canyon Visitor Center
South Rim Trail
Grand Canyon Railway
Village Loop Bypass
Village Loop Drive
S Entrance Rd
Grand Canyon National Park Headquarters
S Entrance Rd
Yavapai
Grand Canyon Village
Tonto St
Kaibab St
Boulder Alley
Boulder St
Center Rd
Sunset Dr
S Entrance Rd

Now what?


Yavapai Point
Hermit Rd
S Entrance Rd
Mather Point
Grand Canyon Visitor Center
South Rim Trail
Angel Trailhead
Canyon Railway
Village Loop Bypass
Village Loop Drive
S Entrance Rd
Grand Canyon National Park Headquarters
Yavapai
Grand Canyon Village
Tonto St
Kaibab St
Maswik Lodge
Boulder Alley
Boulder St
Center Rd
Sunset Dr
Zuni Way
Market Plaza Rd
Coconino St
Muav Ct
Park Cir
Havasupai St
Hualapai Rd
Albright Ave
Pine Loop
Oak Loop
64

Time

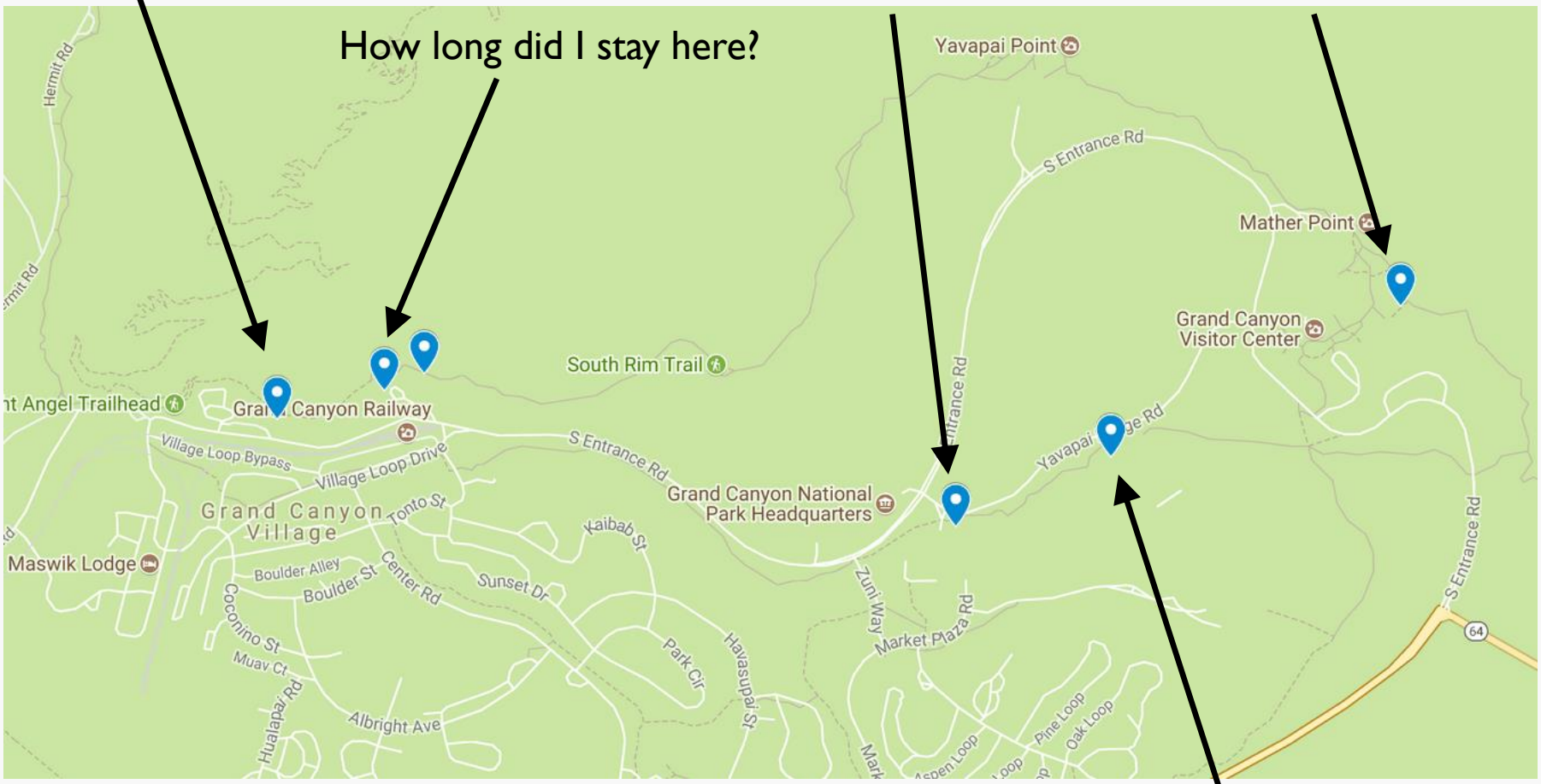
```
<when>2017-03-30T22:16:05Z</when>
<gx:coord>-112.1206089 36.0538447 2110</gx:coord>

<when>2017-03-30T22:15:32Z</when>
<gx:coord>-112.1206895 36.0541252 2108</gx:coord>

<when>2017-03-30T22:14:41Z</when>
<gx:coord>-112.1161455 36.0566548 2117</gx:coord>

<when>2017-03-30T22:13:41Z</when>
<gx:coord>-112.1110006 36.0585582 2123</gx:coord>
```

How often am I here?

How long did I stay here?

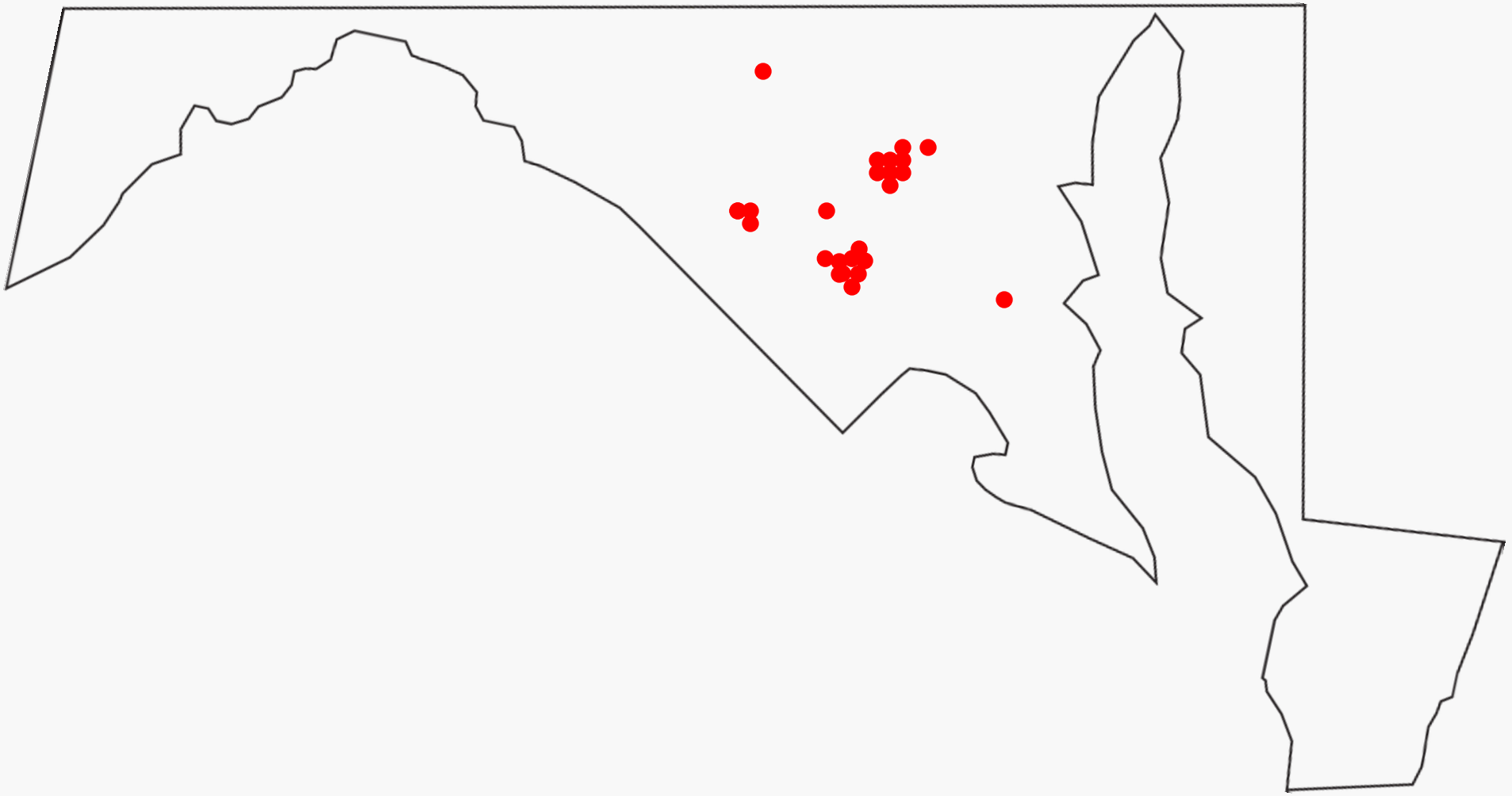
Am I usually here on weekends, weekdays, or both?

When was I last here?

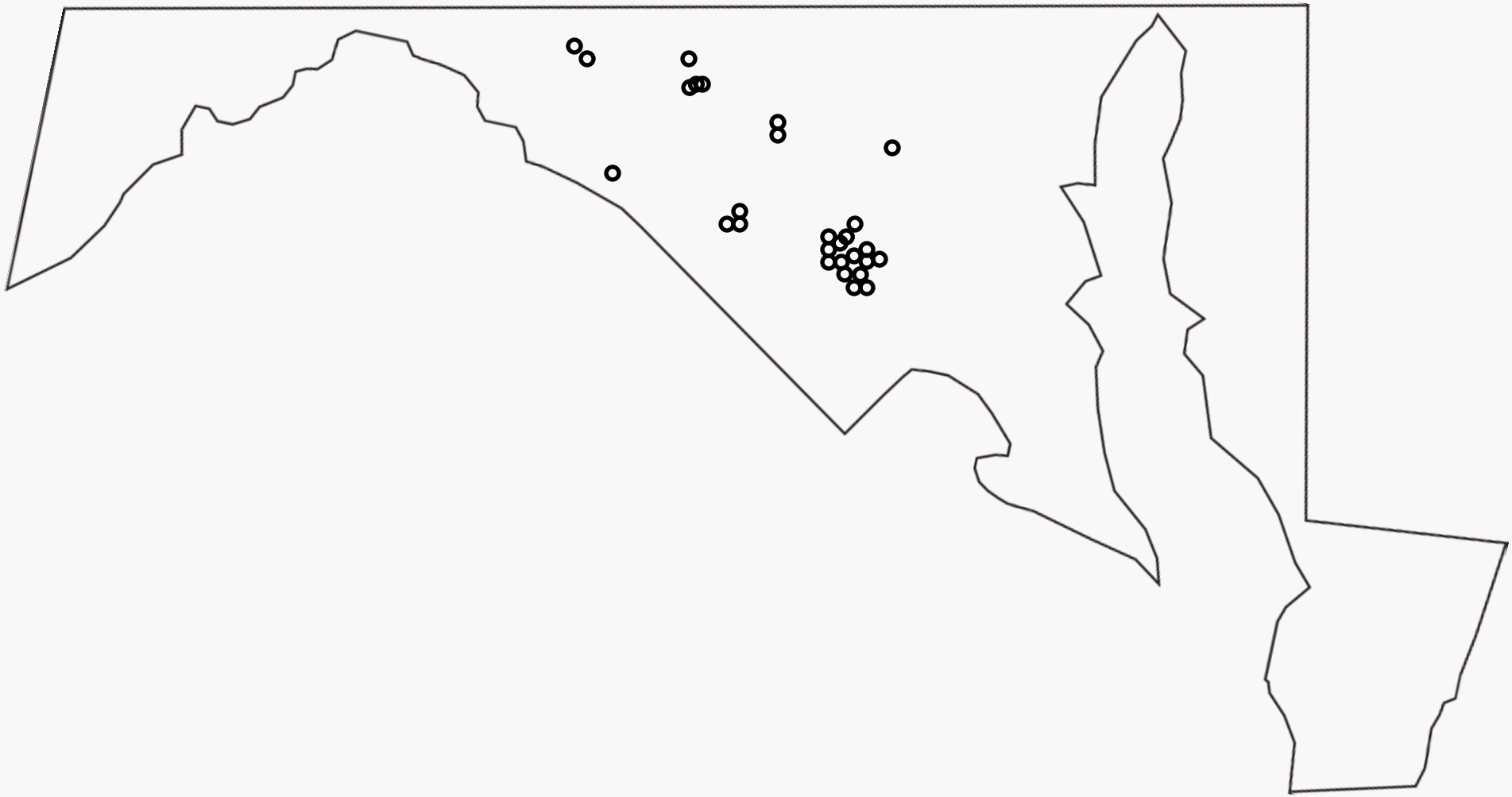


What times of the day can I be found here?

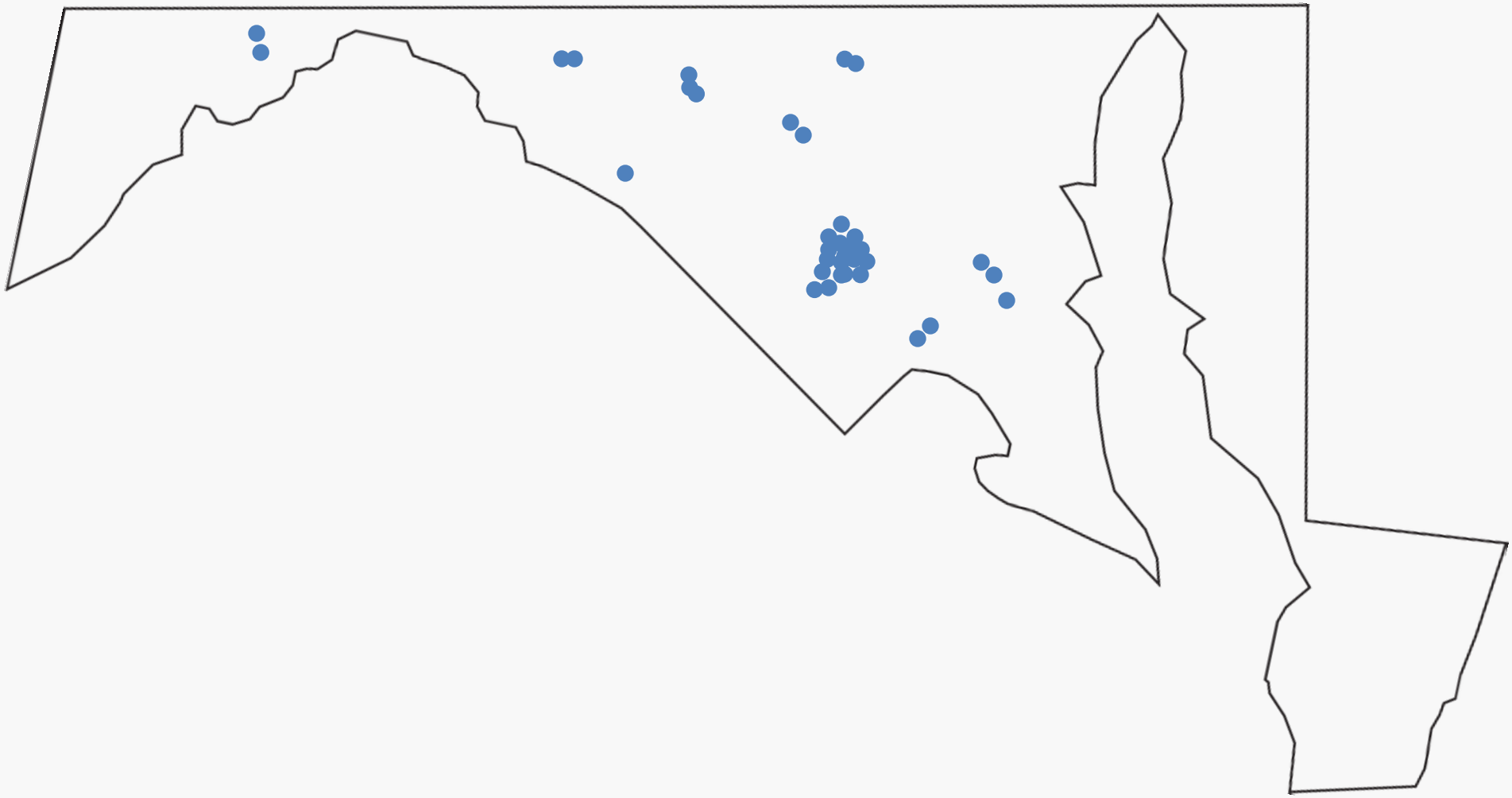
Weekday Day Points



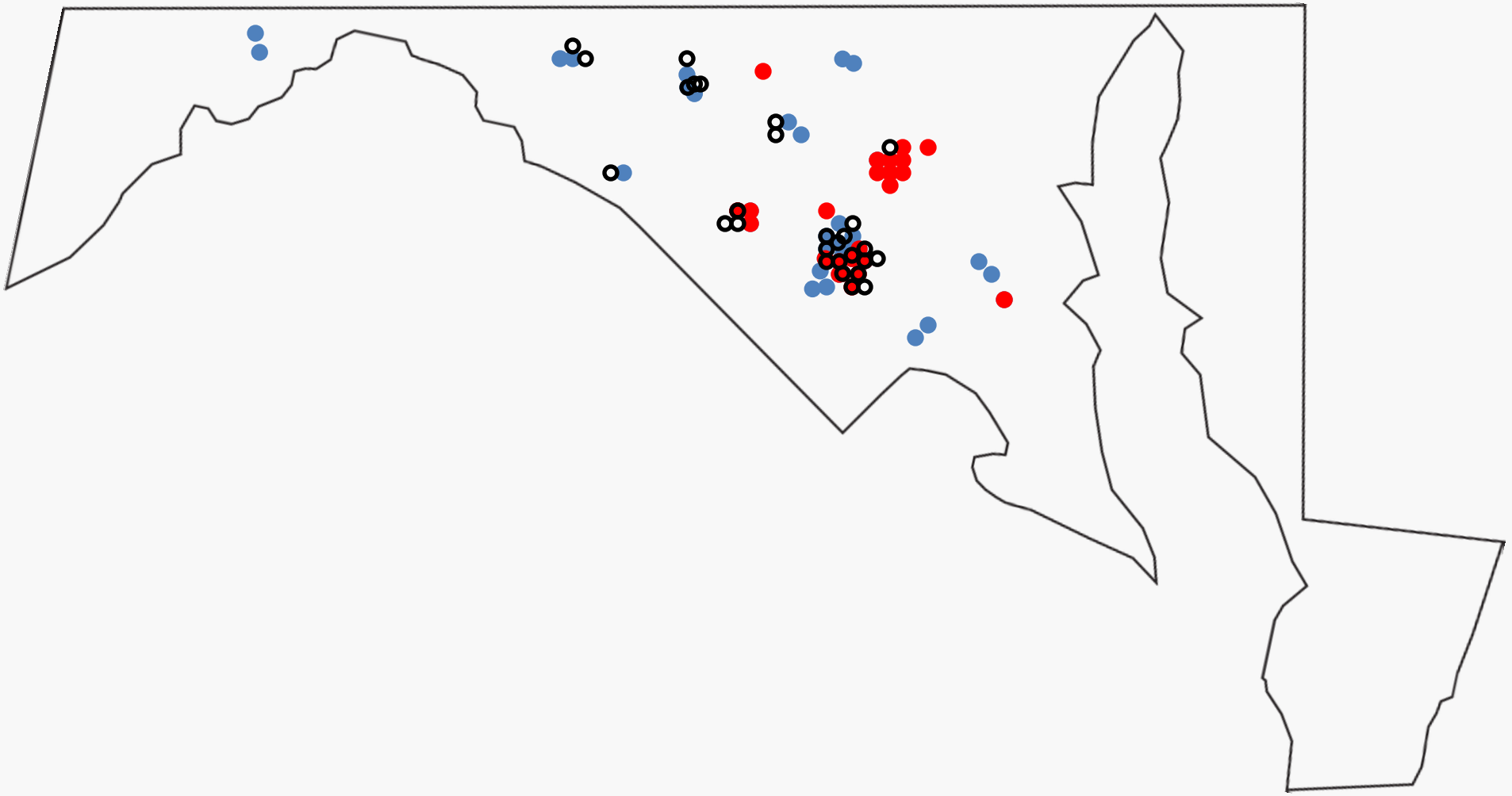
Weekday Evening Points



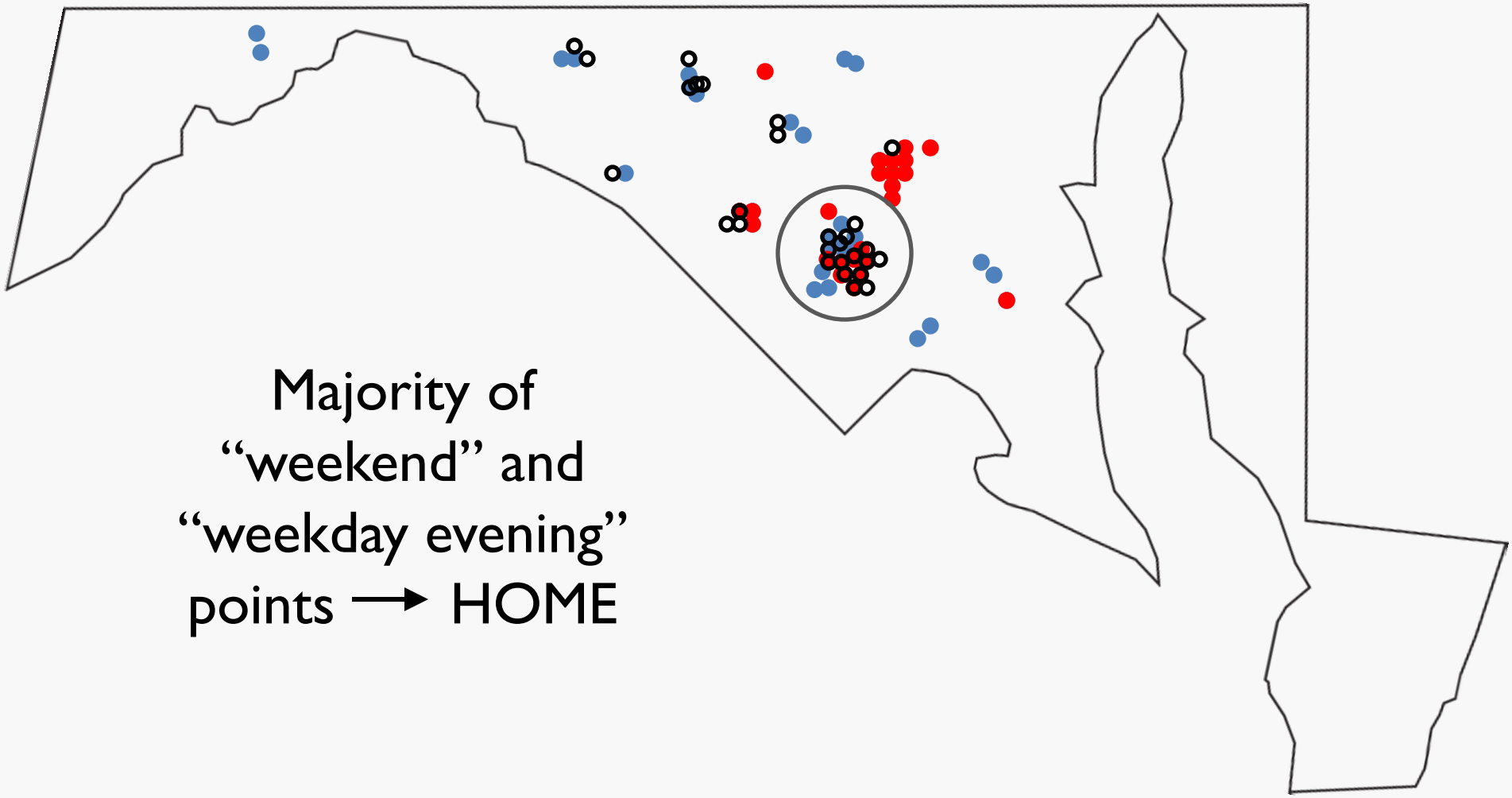
Weekend Points



All Points



All Points



Majority of
"weekend" and
"weekday evening"
points → HOME

If I Know Where "Home" Is, I Can Calculate…

- At a specific time / day of week, probability of being:
  - home
  - away from home

If I Know Where "Home" Is, I Can Calculate…

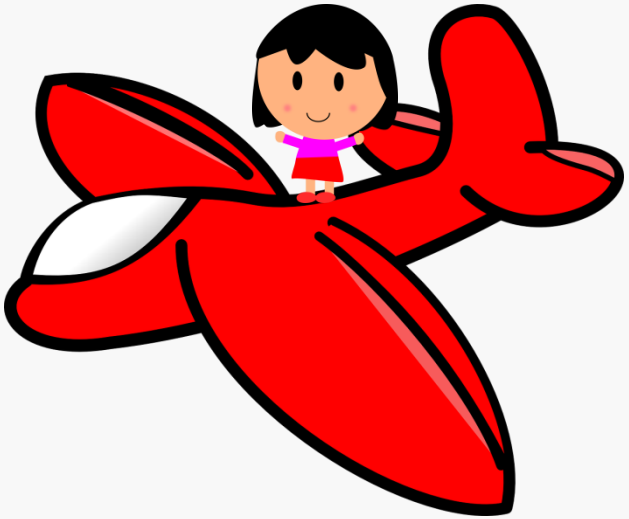
- Average local travel radius
  - Use points when I was home at **both** the beginning/end of the day
- Max local distance I've driven

If I Know Where "Home" Is, I Can Calculate…

- Distance from home for any recorded data point
  - Assume traveling if:
    - Not home that day
    - Home only at beginning/end of day

Significant Location Types Identified

HOME

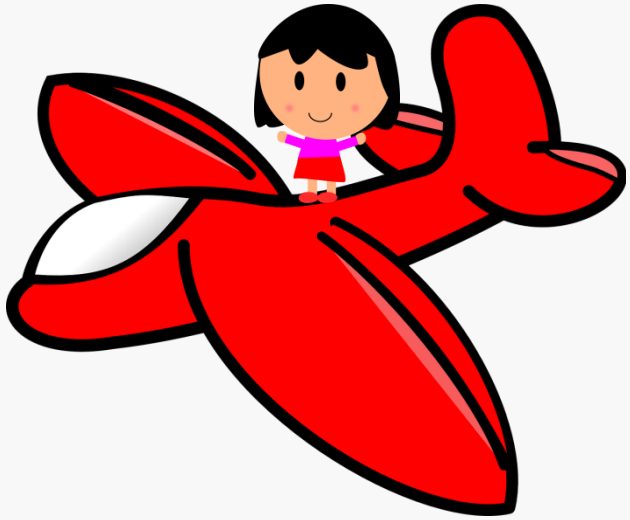


AWAY FROM HOME

LOCAL

Away From Home Locations

- Hotels (Vacation, Conferences)
- Tourist/Shopping Venues



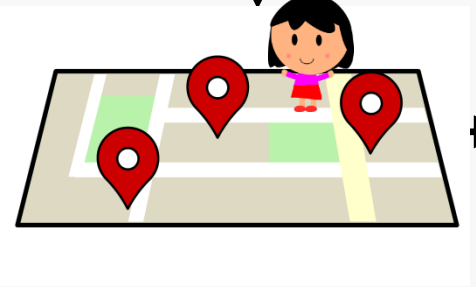
Local Locations

- Favorite Breakfast / Lunch / Dinner Spots
- Grocery Store
- Running Trails

Specific Local Locations

- *Work*: If I'm there on weekdays >= 5 hours
- *Weekend*: If I'm only there on Sat / Sun
  - Concert Venue
- *Same Day*: If I'm only there on a specific day of the week
  - Farmer's Market
  - Trivia Night

Data Science Process



Gather Data

Clean Data

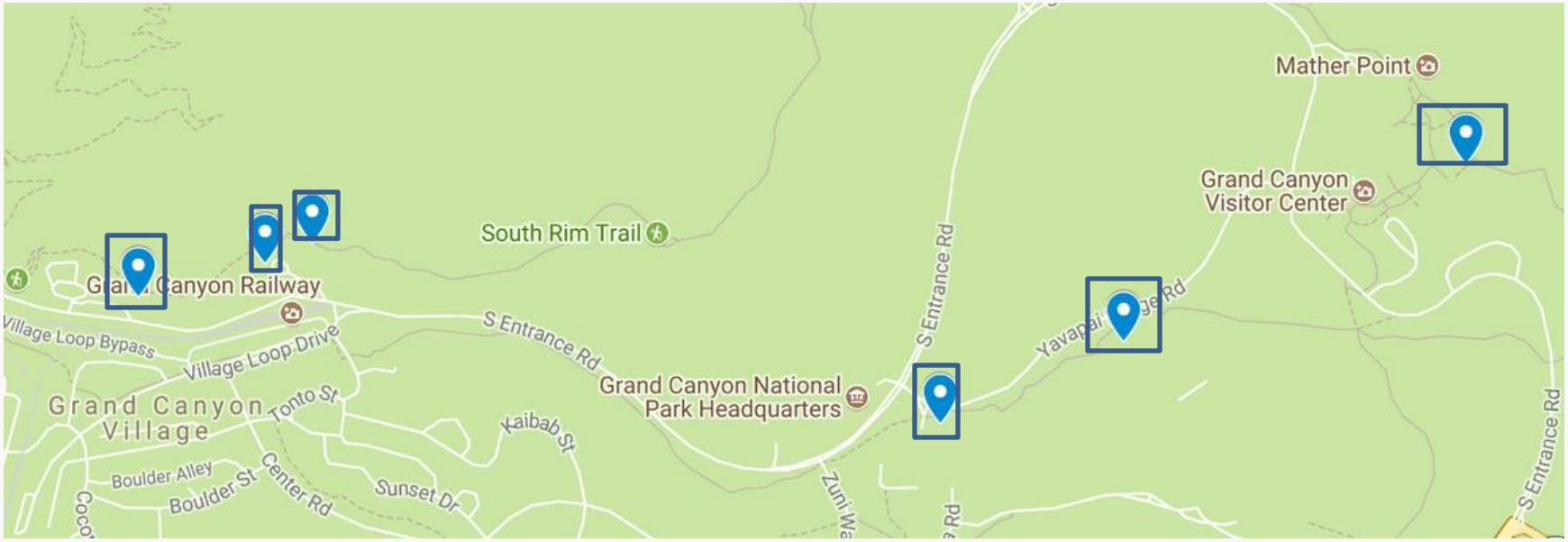
Exploratory Data Analysis

Build Model

Present Findings

Significant Location Details

- Lat / Long Boundaries
- Location Type
  - *AwayFromHome*
  - *Local (home, work, weekend, sameDay)*



Data Point Details

Original

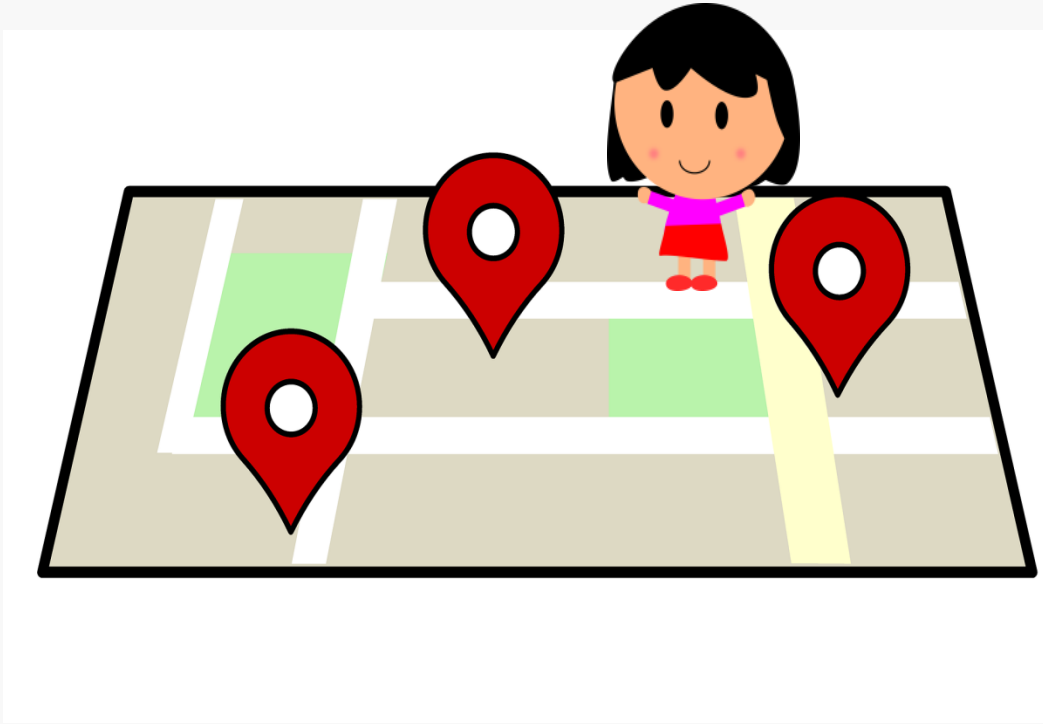
```
<when>2017-03-30T22:16:05Z</when>
<gx:coord>-112.1206089 36.0538447 2110</gx:coord>
```

New

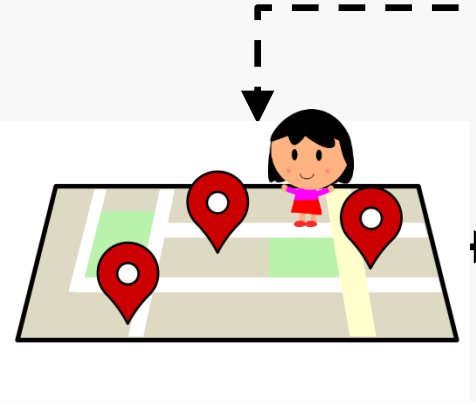
```
<distanceFromHome>1949.46</distanceFromHome>
<locationLabel>cluster3</locationLabel>
<description>awayFromHome</description>
```

Model of Me

- Dates away from home
- Local travel radius
- Likelihood of being at a location by day / time
- Significant locations



Data Science Process



Gather Data

Clean Data

Exploratory Data Analysis

Build Model

Present Findings

DEMO

Questions I can ask the data

- Where was I on August 9, 2017 at 2:18PM ?
- Predict where I will be on Monday at 8:45AM.
- Predict when I am likely to be away on Saturday.
- Predict whether I'll be home on Sunday at 10PM.

Expanded Questions

- How many days was I out of town in July?
- When was I at work on a weekend?
- How many times did I visit the grocery store last month?
- How long does it usually take to drive to work?
- When was I last at the Grand Canyon?

DISCUSSION

Assumptions

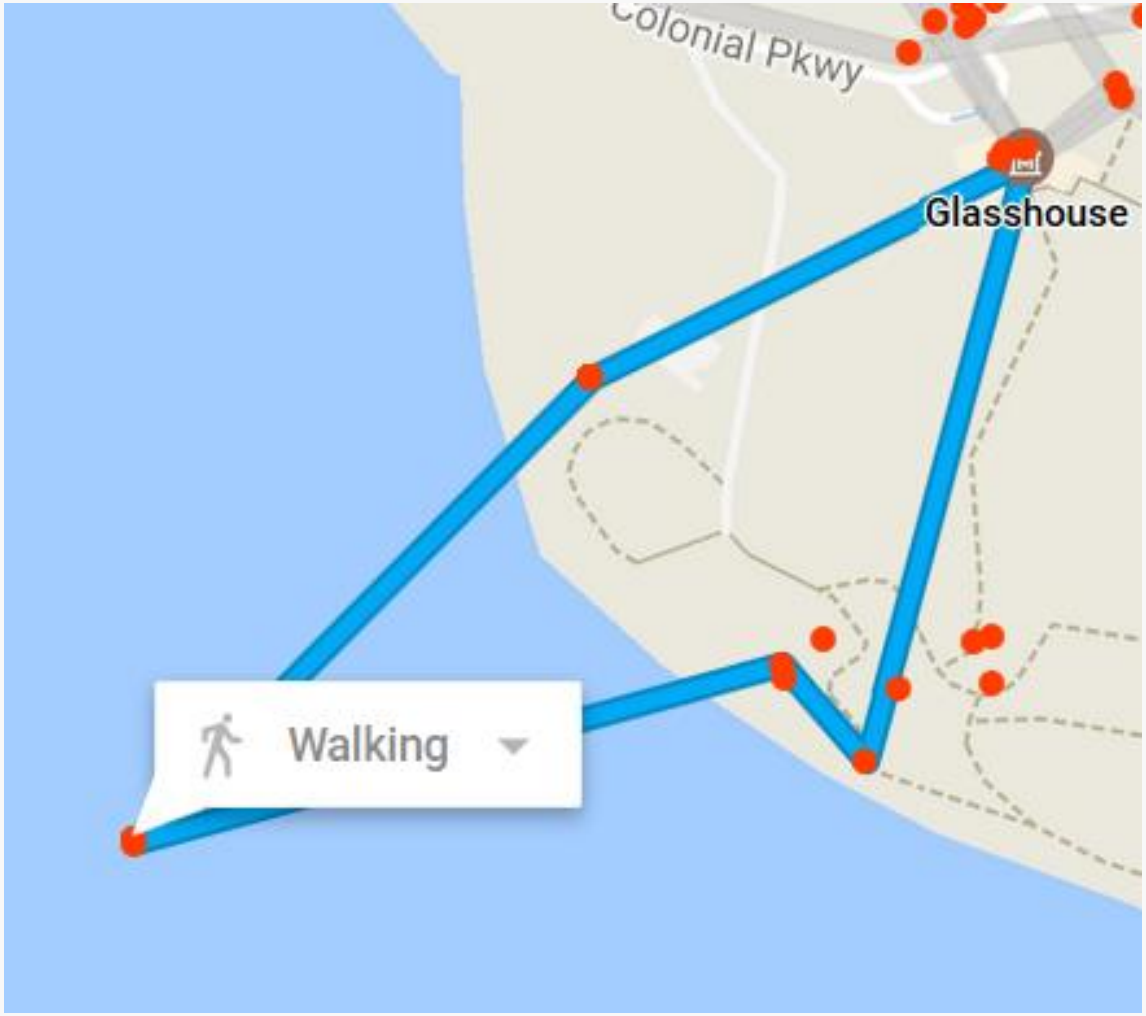
- Regular schedule
- “Normal” work habits

- Home
  - More often than anywhere else
  - More often on weekday evenings & weekends

When This Doesn't Work

- Irregular schedule / lots of travel
- Not enough points
- Bad technology
  - signal
  - hardware

Bad Technology



Cautions

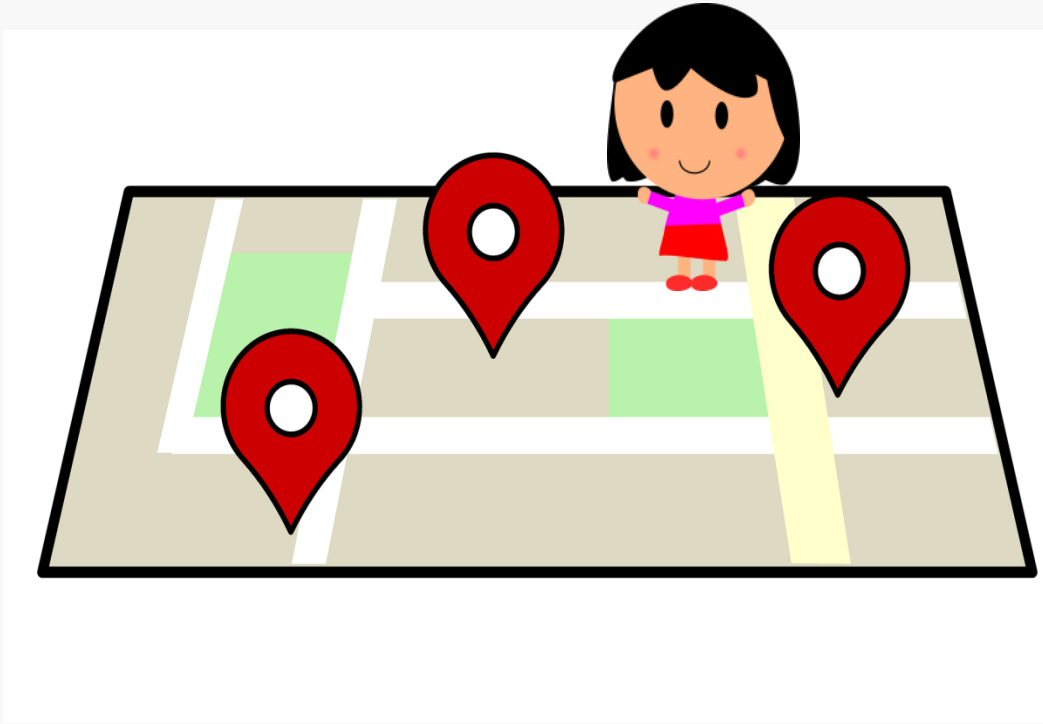
- Analysis is a general pattern of behavior
- Locations may be inaccurate (Google itself asks for corrections)
- A long traffic light can be a “location”

Google already asks for your Home and Work addresses…

…which means that they already know your significant locations!

Would you share this info?

- Dates away from home
- Local travel radius
- Likelihood of being at a location by day / time
- Significant locations



Who could have it

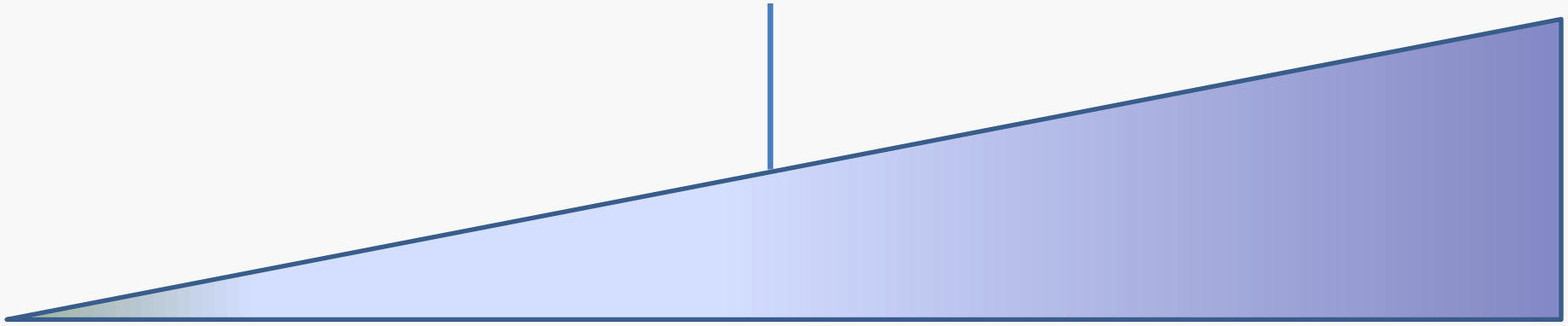
- Products and apps
- Companies that access data
- Companies that buy / share data

Implications

Benign

Implications

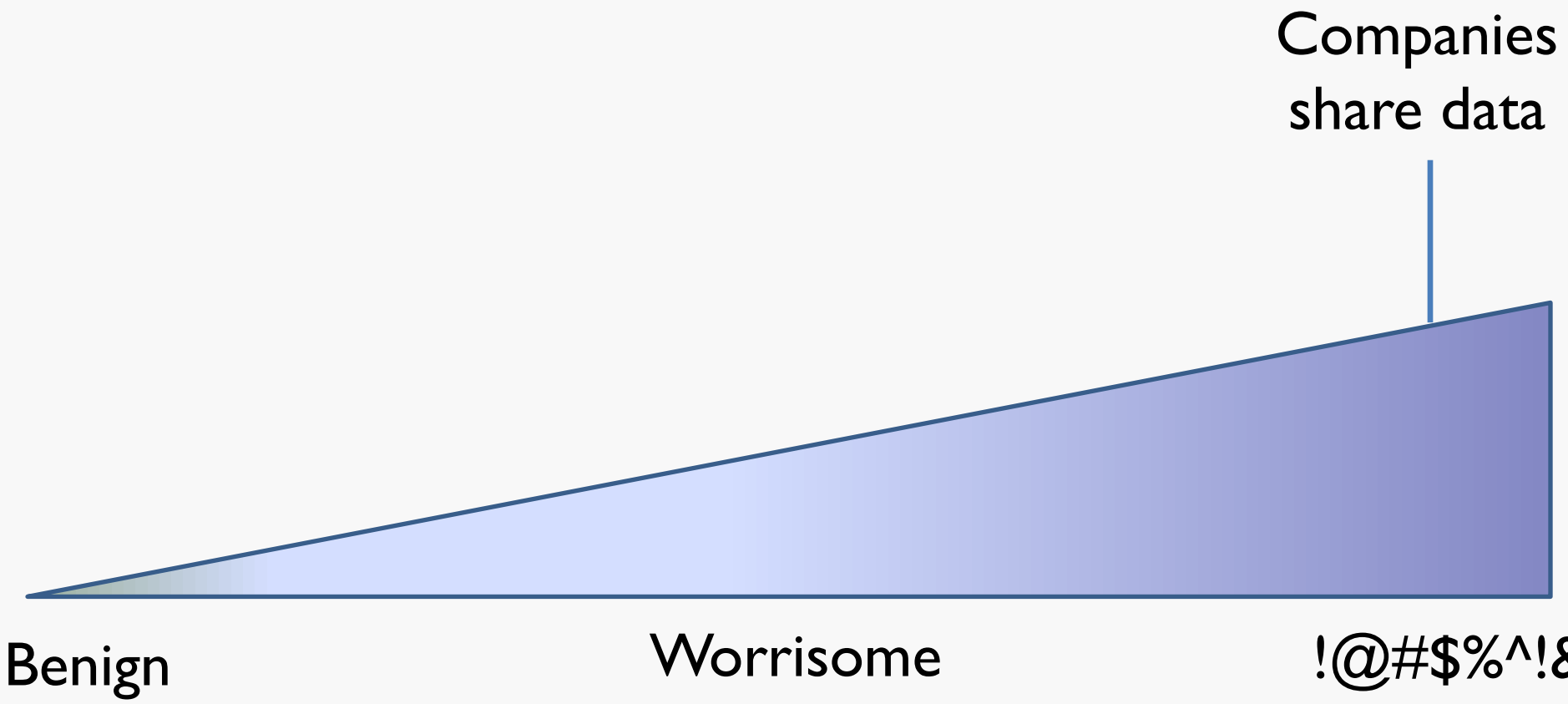
Your insurance claim was denied due to…



Benign

Worrisome

Implications



Your Data, Your Choice

Further Information

- Code (Jupyter Notebook)

  https://github.com/laconicllama

- Contact

  laconicllama@hotmail.com

Questions?