## Fires and Rain in Brazil

## December 2019

### Team Members

Joao Costa

Michelle Fitzpatrick

Stacy Giauque

Hannah Moy

# Project Plan

### Project Overview

Analyze the relationship between annual rainfall and forest fires in Brazil by State.

#### Research Questions

**Step 1: Decompose the “Ask”**

Primary analysis:

* How do precipitation rates vary across Brazil, by state/geography?
* How does the amount of precipitation change, given seasonality?
* How do forest fires vary across Brazil, by state, by state/geography?
* How does the number of forest fires change, given seasonality?

Secondary questions:

* Outlier events, if any? 10-year forest fire or 100 year flood?
* Magnitude and frequency of fire and rainfall events? Increasing vs. decreasing over time.
* How does seasonality correlate to winter/spring (wet season) or summer/fall (dry season)
* Dry Season in Brazil is from March to November
* Wet Season in Brazil is from December to February

#### Data Sources

**Step 2: Identify Data Sources**

Government of Brazil via Kaggle

* Precipitation by month from 1998 to 2017

<https://www.kaggle.com/fabiopotsch/precipitation-in-brazil>

* Forest Fires by month from 1998 to 2017

<https://www.kaggle.com/gustavomodelli/forest-fires-in-brazil/data#_=_>

* Map and definitions of 6 Brazilian geographic biomes

<https://www.arcgis.com/home/item.html?id=54ec099791644be4b273d9d8a853d452>

* Interactive website as benchmark POA for analysis

<http://terrabrasilis.dpi.inpe.br/app/dashboard/deforestation/biomes/legal_amazon/increments>

**3 data sets:**

* monthly precipitation and number of forest fires by
  + month
  + State (25 out of 27 States reported)
  + from 1998 to 2017 (approximately 6,000 data points in both sets)
* map and definitions of Brazil’s 6 geographical biomes
  + Biome borders
    - Amazon, Tropical Savannah, Desert (by vegetation type), Atlantic Forest, Tropical Wetland, Grasslands
* looking at a possible fourth data set which tracks deforestation by year by State by hectare, sourced from Brazilian Government
* 259 hectares per square mile
* Brazil 3.286 million square miles
* http://dados.gov.br/dataset/sistema-nacional-de-informacoes-florestais-snif

#### Task Summary

**Step 3: Define Strategy and Metrics**

* Graph and Table of average annual rainfall per year by State x
* Graph and Table of average annual forest fires per year by State x
* Correlate biomes with Legal Amazon
* Correlate rainfall with Legal Amazon
* Correlate seasonality
* Geographical heatmap for rainfall and forest fires
* Look at correlation between rainfall and forest fires across geography, seasonality, frequency, magnitude of events

Dry Season - March to November

Wet Season - December to February

**Step 4: Scrape Data**

Merge 2 csv data sets

* Forest Fires by state by month by year
* Precipitation by state by month by year

Eliminate NAN’s

* Missing data for unreported states (first pass, there are 2 states missing: Parana, Rio Grande do sul)

**Step 5: Build Retrieval Plan - Pull Data with Python/Yelp API**

* Looking at a possible fourth data set which requires more formatting
* Google Maps API:

https://console.cloud.google.com/google/maps-apis/overview?consoleReturnUrl=https:%2F%2Fcloud.google.com%2Fmaps-platform%2F%3Fapis%3Dmaps,routes,places%26project%3Dmy-project-1574522664929&consoleUI=CLOUD&project=my-project-1574522664929

**Step 6: Retrieve the Data**

* Retrieved data in csv format for 3 data sets

**Step 7: Assemble and Clean Data to make sense**

* Standardize State names
* Standardize date format (DD/MM/YYYY)
* Translate from Portuguese to English

**Step 8: Analyze for Trends**

**Fire**  
Sum by country / year  
- graph visual  
Sum by state / year  
- graph visual

**Rain**  
Sum by country / year  
- graph visual  
Sum by state / year  
- graph visual  
Dig into specific interesting years  
- graph visual

**Combined**  
Fire vs Rain  
- graph visual  
- determine outliers (more fire than rain one month / year? why?)

**Heat map**  
- graph visual (annual average - every 5 years)  
- graph visual (by season throughout 5 years)

**Step 9: Acknowledge Limitations**

* Not all Brazilian States have equal reporting capability, capacity to report
* What’s “In” and “Out” of scope?

**Step 10: Make the Call or Tell the Story and The Insight (and explain why and provide “the INSIGHT.”)**

*Additional Considerations:*

***Data Structure***

* *Make sure you consider the structure of your data before you begin to transform it. Check your data types, the number of columns, and do a visual “sanity check” to make sure the data you downloaded seems accurate.*
* *Data can be on different levels of granularity. When combining datasets ensure that you are doing so on the same level of data. Examples of different levels of data are:*
  + *Country*
  + *Region*
  + *State/Province*
  + *City*
  + *Individual People*
* *When combining data, ensure that you have a unique key that you can create or utilize to combine the datasets together.*

***Research***

* *Has your subject matter research indicated common themes across datasets?*
* *Can these themes be represented by available variables in the datasets?*
* *Are there metrics that can be identified across multiple datasets that could enrich/supplement your analysis?*

***Analysis***

* *Will you rely on quantitative or qualitative data for your analysis, or both? Your choice of metrics will determine the kind of analysis you will be able to do, as these data behave differently in different analyses. Make sure you understand how your analysis tool of choice treats these variables, and if there are any options available to compare quantitative data (i.e. temperature) to qualitative data (i.e. survey answers on how people feel about certain temperatures). Research methods like dummy variables, that can be used to turn qualitative data into categorical data that allows the comparison to be made.*

***Insights***

* *What patterns are emerging in the data? an “old” view of biosphere and a “modern” view after 2004 Government of Brazil’s ecological initiatives*
* *How will you determine if these patterns are related to each other alone, or if they are an indication of another variable you may not have yet considered (this is where subject matter research be used as a guide) succinct use of python data analysis/statistical package: correlation coefficient, ANOVA*

***Key Take-Aways***

* *What story does the data tell? significant inverse relationship between rainfall and forestfires but deforestation is more closely linked to “Biome” and “other factors” including likely farming*
* *What are the most important components and/or actionable insights you can derive from the data? visualization of trends, correlations between variables*
* *What is the best way to represent the data (and story) to your audience? Powerpoint, graphs and tables*

#### *Recommended: Schedule*

* *How will you know you are on track? stay succinct, assign deliverables and timelines*
* *Who is responsible for which components of the projects? Hannah (rainfall,) Michelle (forestfires,) Joao (deforestation,) Stacy (roadmapping/Visual API mapping/notion of Legal Amazon)*
* *What is in and out of scope? in-scope: simple trend analyses of data, correlations*

*out-of-scope: Government of Brazil, UN, Lobbyists’, Environmentalists’ initiatives*