Names: Daniel Luft-Martinez, Hoa Nguyen

Title:

Our Kleptocracy  
 -  
Delving into the panama/paradise papers

Questions:

On a global scale:

Who?  
 Which individuals/entities/groups have been named in the leak?  
 Can we link them with their shell companies?  
 Where are they?

What is their status?

politically?

economically?

affiliations?

Can we use patterns in the data to further identify others?

Data Sources:

large csv’s formatted for the neo4j data base

<https://offshoreleaks.icij.org/pages/database>

Wikipedia Pages

<https://en.wikipedia.org/wiki/Panama_Papers>

<https://en.wikipedia.org/wiki/Paradise_Papers>

Code References:

<https://www.programcreek.com/python/example/88086/py2neo.Graph>

<https://medium.com/neo4j/py2neo-v4-2bedc8afef2>

<https://plot.ly/python/mapbox-county-choropleth/>

<https://plot.ly/python/bubble-maps/>

Visual References:

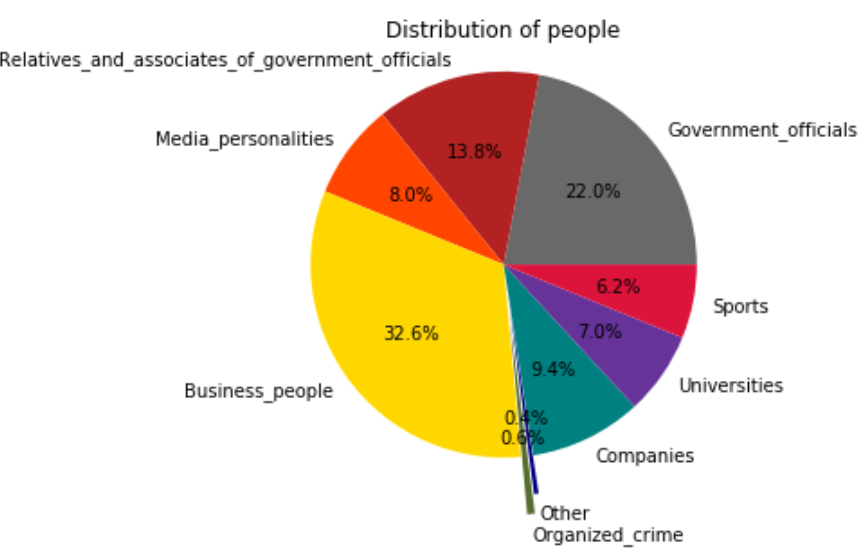
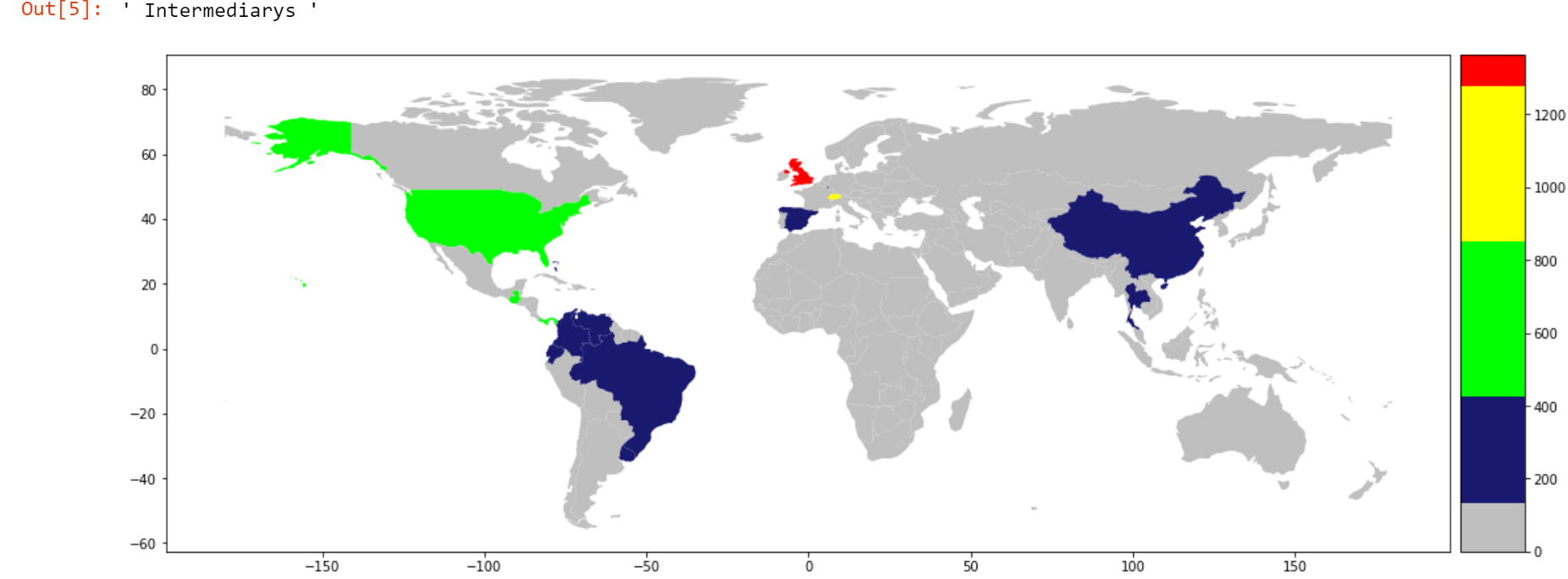
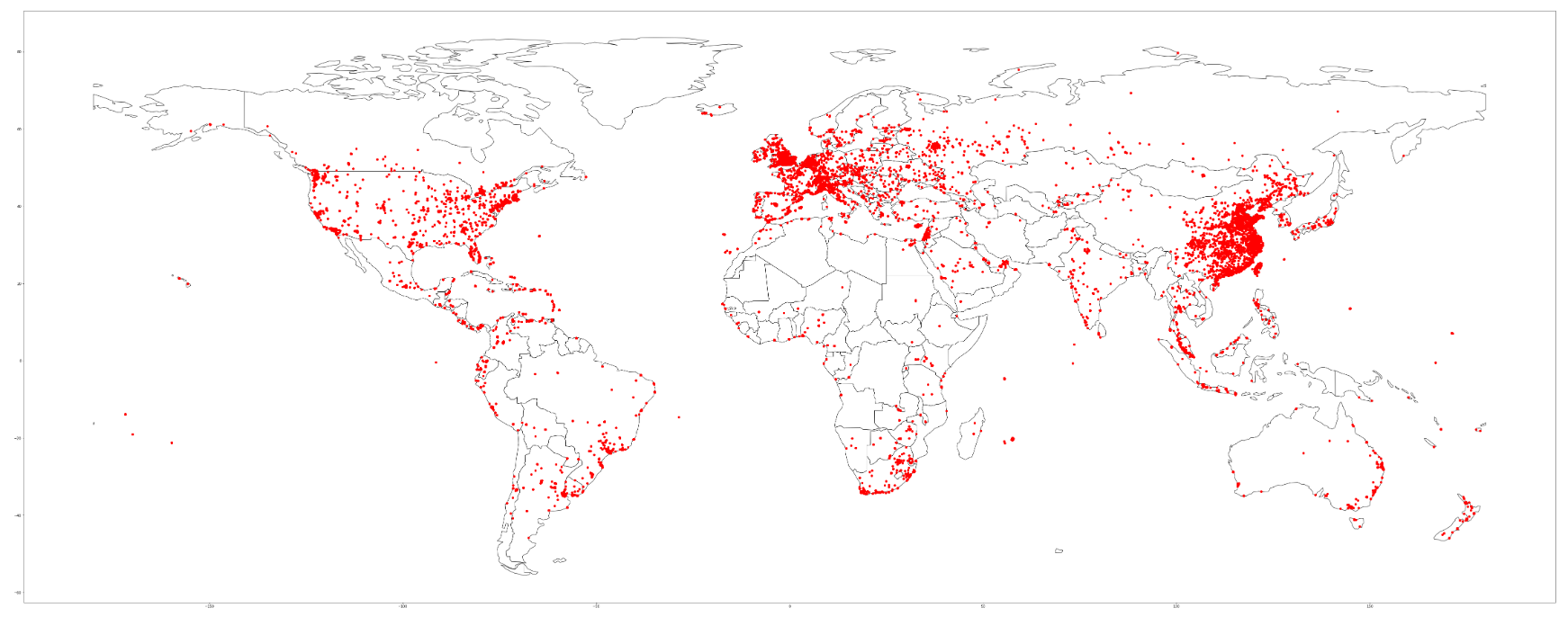
<http://infographic.statista.com/normal/chartoftheday_4606_the_global_extent_of_the_panama_papers_leak_n.jpg>

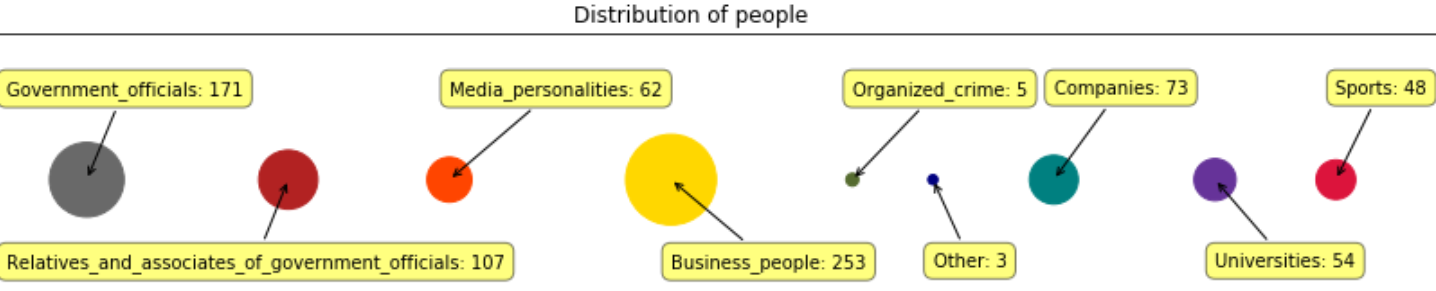
<https://finnaarupnielsen.files.wordpress.com/2016/05/panamapapersoccupations.png?w=705>

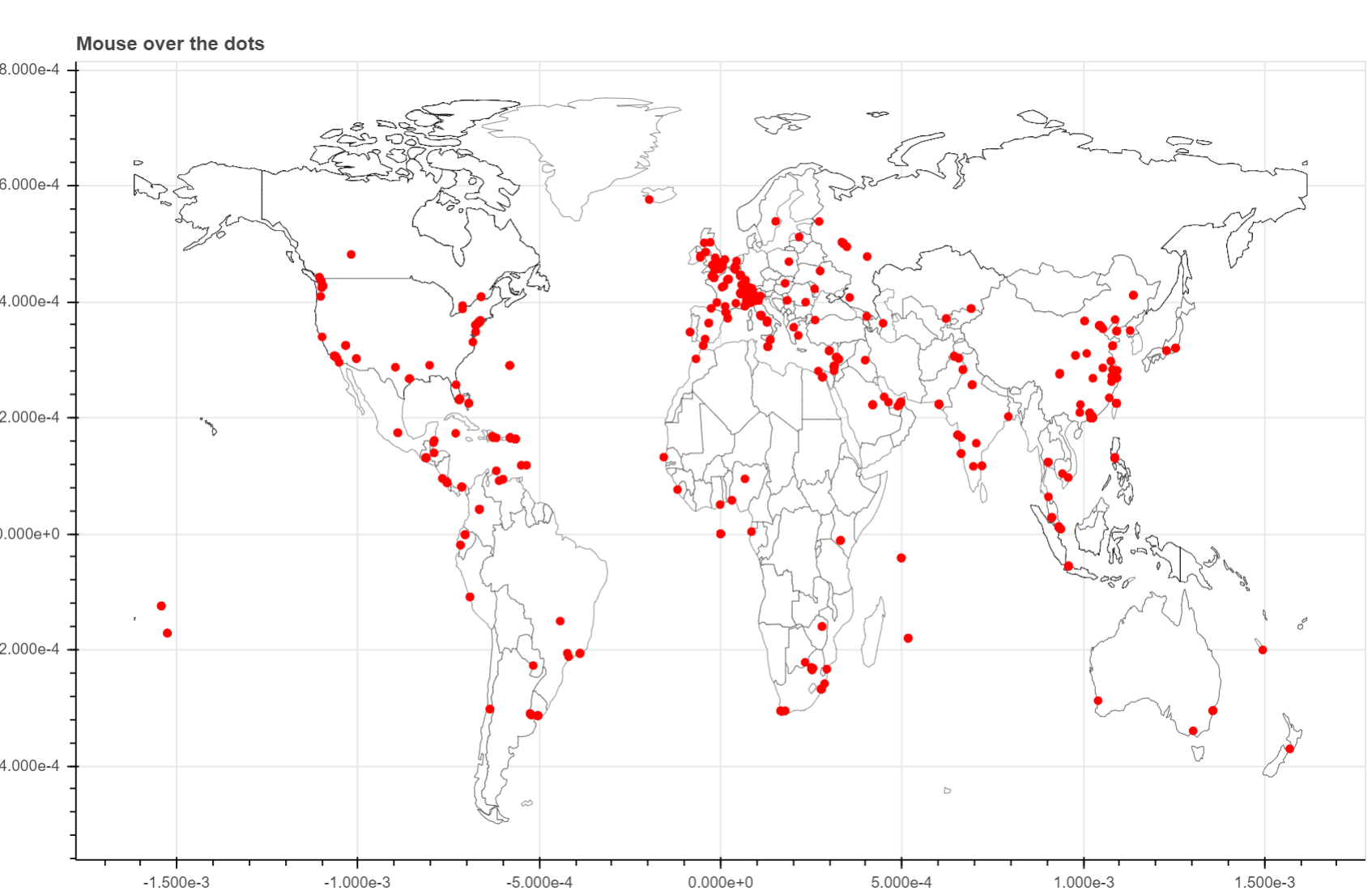
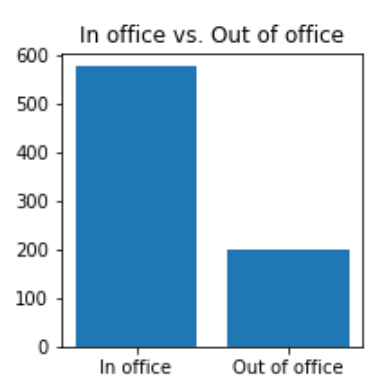
[https://cdn-images-1.medium.com/max/2000/1\*EBHHwzYnGQXYqzYZGrmj2g.png](https://cdn-images-1.medium.com/max/2000/1*EBHHwzYnGQXYqzYZGrmj2g.png)

<https://www.icij.org/investigations/paradise-papers/explore-politicians-paradise-papers/>

Results:







Libraries Used:

pandas

numpy

string

fuzzywuzzy

random

sys

os

json

joblib

matplotlib

mpl\_toolkits

geopy

geopandas

bokeh

pandas\_bokeh

pyproj

pygeocoder

PIL

requests

io BytesIO

urllib3

BeautifulSoup

wikipedia

google\_images\_download

requests.packages.urllib3.exceptions InsecureRequestWarning

List of data wrangling/analysis/visualization tasks accomplished:

Daniel:

Scraping/cleaning/merging separated data from primary CSVs

Geolocation (extremely long processing time 2-6 requests/second for about 24 hours)

Matplotlib mapping

Fuzzy Matching (where most of the misnames in final viz happen)

Bokeh mapping/plotting/interactions (extremely time consuming to figure out)

Hoa:

Scraped wikipedia for links and images

scraped google for images

cleaned wikipedia data

visualized data from wikipedia

IPYNB descriptions:

Daniel: daniel\_luft-martinez\_

1 - a first look at the primary data source

2 - color mapping of registered address country codes

3 - geolocation for addresses

4 ,5 - bokeh testing ground

6 - static pin map of all addresses in primary data set, fuzzy matching to Hoa data, and the final Bokeh interactive map of the “interesting” people that were fuzzy matched

7 – Final presentation ipynb, all visualizations, pull from joblib files to reduce runtime when possible

Hoa: hoa\_nguyen\_

1 - ''' scrapes wikipedia for panama and paradise entities and extracts their website link and image, stores data in a json file, contains some visualizations'''