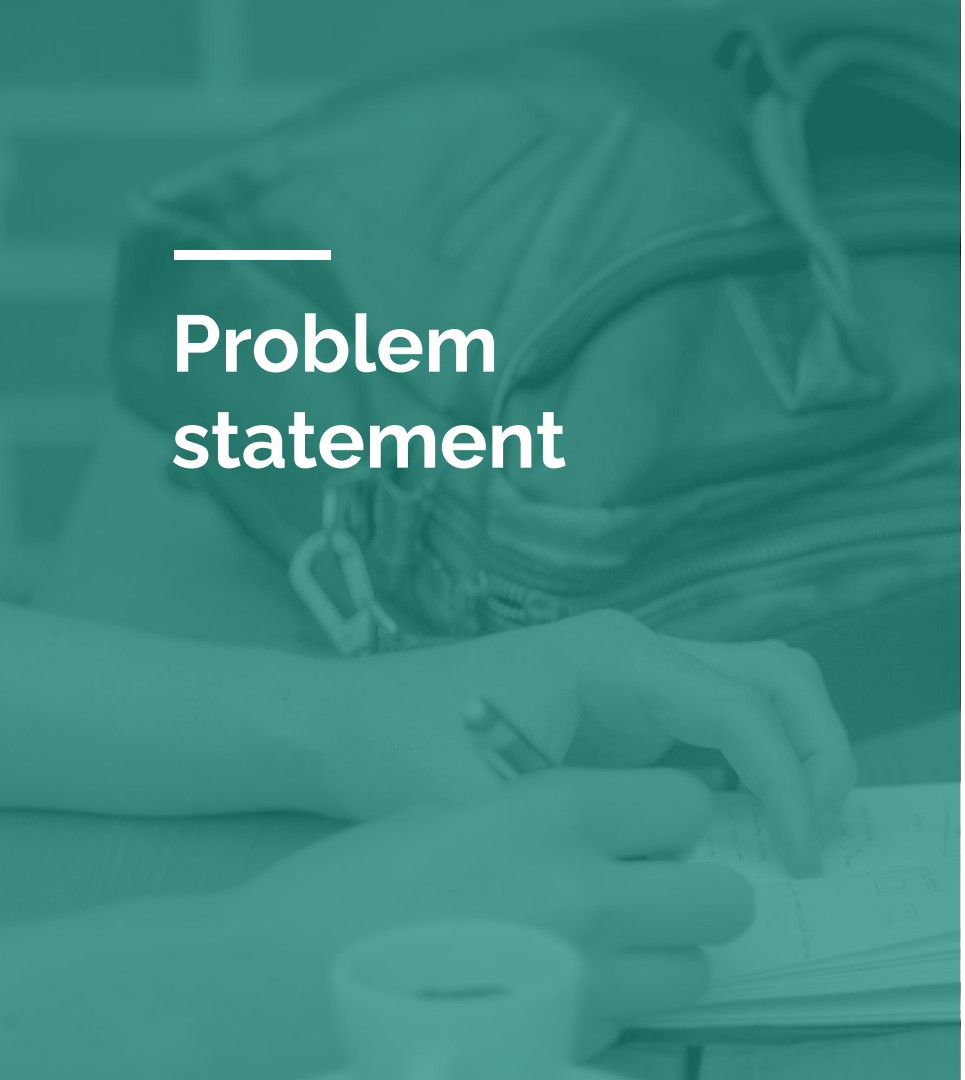
Who Are We?

The Demographics of
US Amateur Radio Licensees
26 January 2022

Outline

The Problem



Problem statement

We do not have demographic data and information for US amateur radio licensees.

“You can’t manage what you can’t measure.”

You can't know whether or not you are successful unless success is defined and tracked. You can’t improve what you can’t (or won’t) measure.

-Peter Drucker

What have we done so far?

Mostly nothing.

We have the impression that amateur radio in the US is almost completely dominated by older white men.

For a telecommunications service that is supposed to be accessible to, and enhance the communications skills of, the general public, not really knowing “who we are” is problematic. Are our impressions wrong or right?

The pain points are numerous. The lack of diverse participation in a fun and easy technical entry path objectively harms our technical workforce readiness.

In the era of STEM, exclusion has real economic and national security implications.

Prior Work

01

There is a good study from 2005 about gender.

(2005 by Ken Harker, WM5R)

A Look at the Numbers

For all amateurs in the FCC database (with a total of 847,809 hams in the June 2004 database I used), 115,266 were categorized as Female (13.60%) 660,798 were categorized as Male (77.94%) and 71,744 had first names that led to a classification of Uncertain (8.46%). If we look at the adjusted percentages, my estimate is that the Amateur Radio population in the United States is 14.85% female and 85.15% male. This is a slightly higher female percentage than the ARRL surveys indicate, and almost three times as high as the percentage of ARRL members who are women.

Prior Work

02

Kai Siwiak, KE4PT is the editor of QEX. His editorial from May-June 2018 issue of QEX has key demographic information about QEX subscribers.

http://www.arrl.org/files/file/QEX_Next_Issue/May-Jun2018/Perspectives.pdf

Mentors

My Amateur Radio mantra is, "Please do your part to lower the average age of hams: Elmer youngsters!" Elmer was personified by David P. Newkirk, W9BRD, in his "How's DX" *QST* column of March, 1971, as a local ham who, "though busy with his own operating, building, arduous studies, chronic family illness, and full social calendar ... miraculously found time to be the big brother to any local youngster or oldster groping uncertainly towards hamdom." Elmer is both a proper noun and an action verb, but the term itself is anachronistic jargon. Times change, and we now prefer *Mentor*!

Consider this. The *QEX* family of hams is aging dramatically. A very recent survey of *QEX* subscribers reveals that about 90% of you, dear readers, are older than 61 years of age, and half of those are over 70. An additional 9% are between 51 and 60 years of age. More than 95% have been licensed for more than 30 years, and nearly two-thirds are retired. What, then, of the future of *QEX*? We need an influx of younger blood — especially between the ages of 40 and 60 — to grow our journal into the near future.

The *QEX* mission is to provide a medium for the exchange of ideas and information among Amateur Radio experimenters, to document advanced technical work in the Amateur Radio field, and to support efforts in advancing the state of the Amateur Radio art. When most of you were at the onset of your ham careers, your main search parameter was "Radio" in your neighborhood library. *QEX* was in its infancy. In contrast, today's generation is immersed in an internet and social media culture. But, you won't find *QEX* there — yet!

Many of your clubs already sponsor ham radio classes and volunteer examiner testing. Please don't stop there. Actively reach out and take the next step. Mentor your newly licensed hams into our fold as *radio-active* hams who can help fulfill the *QEX* mission.

We'd like to hear from you (**qex@arrl.org**), the sages, elders and mentors of Amateur Radio. How can we make the *QEX* mission relevant to the next generations? What can we do to lower the average age of hams? And especially, what can we do to lower the average age of the *QEX* readership?

1% of QEX readers are younger than 51.

Note the acknowledgement that the term "Elmer" is problematic, and the call to action.

Prior Work

03

There is a good study from 2021 about demographics in CW contesting.

By Frank K4FMH
“https://www.amateurradio.com/the-secret-storm-approaching-cw-contesting/”

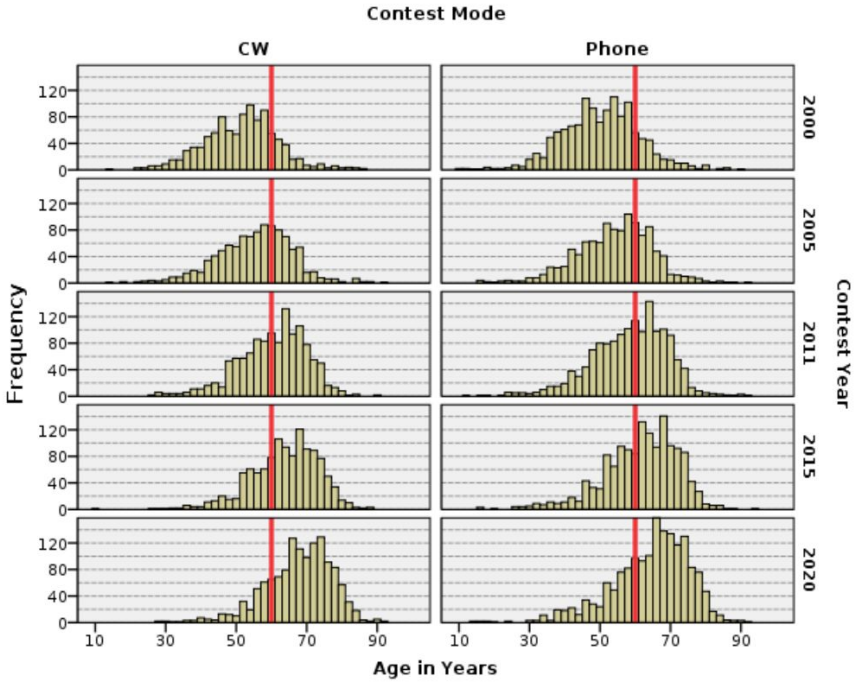


Figure 3. Age Distribution of Sweepstakes Participants by Year and Mode

Prior Work

04

The Royal Society of Great Britain has information about the demographics of people taking license exams in Great Britain.

https://thersgb.org/publications/committees/esc/2021/211203_esc_annual_report%202021.pdf

The 40 – 59 age range accounted for just over 53% (46% 2019) of candidates taking the exams (Table 3). The number of youngsters (<=19 years old) taking the exams, 268, fell in 2020 compared to 2019 (400), despite the increase in overall candidate numbers. Possible reasons for this is are the disruption to school education during the pandemic, requiring more effort from them to keep up, and the reluctance to engage in yet more online learning.

The % female to male candidates remains disappointingly poor at less than 5%. Of the 3,087 candidates taking the Foundation exam in 2020 only 172 were female, with 22 out of 868 at Intermediate and more alarmingly only 6 out 38 at Full.

Supporting information

05

Do we need to be convinced that diverse teams and populations outperform homogenous ones?

Lucky for us, there's a pile of science.

Good survey article:
https://hbr.org/2016/11/why-diverse-teams-are-smarter

Better R&D Innovation

06

“Our findings support the assertion that gender diversity within R&D teams generates certain dynamics that foster novel solutions leading to radical innovation. The results indicate that gender diversity is positively related to radical innovation. However it does not promote incremental innovation in the same way.”

Diaz-Garcia, Cristina, et al. "Gender diversity within R&D teams: its impact on radicalness of innovation." Innovation: Management, Policy, & Practice, vol. 15, no. 2, June 2013, pp. 149+. Gale Academic OneFile, link.gale.com/apps/doc/A337288505/AONE?u=anon~207e97f8&sid=googleScholar&xid=9681e325. Accessed 26 Dec. 2021.

Solution Proposal

Solution description

With basic modeling, demographic data about US amateur radio licensees can be constructed.

Gender can be guessed based on machine learning models that assign gender to names.
The ratio of men to women can be estimated, acknowledging that some names will be unknown (e.g. Pat, Leslie). The databases used are geographically focused. For example, Japanese names in the US FCC database are not categorized very well. This can be done manually, or by re-running Japanese names using a Japanese ML database.

Race can be guessed based on probabilistic models that assign a race to a licensee based on census results per zip code or geotagging. Taking into account techniques from Model Thinking, we can refine this probabilistic model to include the effects of *sorting* and *peer effects*. University of Michigan has the most public-facing presentation of this work.

Solution description

With basic modeling, demographic data about US amateur radio licensees can be constructed.

Age is more difficult to model because it has not been collected for US licensees for quite a while, but there are some things that can be done. Look at the CW contesting work.

Thank you to Marty Woll and Dick Norton for sharing a database of age information about amateur radio licensees up until the point it was no longer collected, and for providing a path forward for getting current demographics with respect to age. Excellent suggestions, support, and advice.

Knowing is Better than Not Knowing

Amateur radio licensees are allowed the use of frequency spectrum because it is considered to provide positive **public** benefit.

The amateur service in the US is self-regulated. We need to ensure as **equal access** as possible. **In practical terms, the responsibility for being of service to the general public is on us, not the FCC.**

If we don't know our own demographics, we cannot authentically claim amateur radio is accessible to, or of benefit to, the general public.

FCC really could and probably should be providing demographic information, like it does for professional broadcasting licenses. An organization like ARRL could take this on.

Knowing removes any excuses for not dealing with it. This can be stressful and difficult work.

The Code

https://raw.githubusercontent.com/Abraxas3d/Demographics/master/main.py

```
import os
import random
import webbrowser

import folium
import pandas as pd
from gender_detector.gender_detector import GenderDetector
from uszipcode import SearchEngine

active_licenses = set()
```

```
try:
  with open('C:/Users/Kindl/OneDrive/Documents/Amateur-Radio-Demographics/HD.dat', 'r') as headers:
    # pass
    for active_line in headers:
      active_list: list[str] = active_line.split("|")  # bust up the line at the | symbols, makes a list

      if (active_list[5] == "A"):  # active licenses
        active_licenses.add(active_list[4])

except FileNotFoundError:
  pass


# print(active_licenses)

# we now have a set of active licenses
```

```
HD|12996203|0009826477||AA8NC|E|HA||||||||||N|||||||||N||Foo|D|Foo|832|||||||||12/28/2021||||||N|||||||
HD|13083160|0009806437||KI4ROT|C|HA||||||||||Y|||||||||N||Boo|P|Boo|||||||||12/28/2021||||||N|||||||
HD|13089140|0009844648||N0CAL|A|HV||||||||||N|||||||||N||Dude||Salzberg||self|||||||12/28/2021||||||N|||||||
HD|13089754|0009822626||NI7BAR|A|HV||||||||||N|||||||||N||Eric|R|Smith|||||||||12/28/2021||||||N|N|||||||
```

```
detector = GenderDetector('us')  # It can also be ar, uk, uy.

search = SearchEngine(simple_zipcode=False)  # zipcode demographics lookup

unknown_names =
open(r"C:/Users/Kindl/OneDrive/Documents/Amateur-Radio-Demographics/unknown_names.dat", "w+")

try:
   with open('C:/Users/Kindl/OneDrive/Documents/Amateur-Radio-Demographics/EN.dat', 'r') as my_file:
      # pass
```

EN|13107733|||**AA8NC**|L|L00555463|Loo, Joo A|Joo|A|Loo|||||6515 Someplace LN|Spring|TX|77379|||000|0006435507|I|||||||
EN|14306210|||**KI4ROT**|L|L07755131|Foo, Bar A|Bar|A|Foo|||||999 Cantfind Rd|Smithville|NC|27606||Bar Foo|000|0031757559|I|||||||
EN|91105517|||**N0CAL**|L|L02345348|Cable, Guy D|Guy|D|Cable|||||3245 S Barnards AVE|Nowhere|NC|27244|||000|0028987641|I|||||||

```
male_count = 0
female_count = 0
other_count = 0
unknown_count = 0
punch_count = 0
white_count = 0
black_count = 0
american_indian_count = 0
asian_count = 0
hawaiian_count = 0
other_race_count = 0
two_or_more_count = 0
prob_white = 0.0
prob_black = 0.0
prob_american_indian = 0.0
prob_asian = 0.0
prob_hawaiian = 0.0
prob_other_race = 0.0
prob_two_or_more = 0.0
zipcode_fail = 0
total_count = 0
heat_map_dataframe = pd.DataFrame()
heat_map_dictionary = {}
```

```
    for my_line in my_file:
        my_list: list[str] = my_line.split("|")  # bust up the line at the | symbols, makes a list

        if (my_list[4] in active_licenses) and (my_list[5] == "L") and (my_list[23] == "I"):

# active valid individual license? continue
```

EN|13107733|||AA8NC|L|L00555463|Loo, Joo A|Joo|A|Loo|||||6515 Someplace LN|Spring|TX|77379|||000|0006435507|I|||||||
EN|14306210|||KI4ROT|L|L07755131|Foo, Bar A|Bar|A|Foo|||||999 Cantfind Rd|Smithville|NC|27606||Bar Foo|000|0031757559|I|||||||
EN|91105517|||N0CAL|L|L02345348|Cable, Guy D|Guy|D|Cable|||||3245 S Barnards AVE|Nowhere|NC|27244|||000|0028987641|I|||||||

```
my_zipcode = search.by_zipcode(my_list[18][:5])

#    [{'key': 'Data', 'values': [{'x': 'White', 'y': 10439}, {'x': 'Black Or African American', 'y': 46},
#                    {'x': 'American Indian Or Alaskan Native', 'y': 60},
#                    {'x': 'Asian', 'y': 93},
#                    {'x': 'Native Hawaiian & Other Pacific Islander', 'y': 2},
#                    {'x': 'Other Race', 'y': 46}, {'x': 'Two Or More Races', 'y': 154}]}]

if (my_zipcode.population_by_race == None):
   zipcode_fail += 1
   #print("-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=")
   #print("Zipcode access failed.")
   #print("zip code is:", my_list[18][:5])
   #print(my_line)
   #print(my_zipcode)
   #print("-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=")
```

else:

```
 prob_white = my_zipcode.population_by_race[0]['values'][0]['y'] / my_zipcode.population
# print(prob_white)
prob_black = my_zipcode.population_by_race[0]['values'][1]['y'] / my_zipcode.population
# print(prob_black)
prob_american_indian = my_zipcode.population_by_race[0]['values'][2]['y'] / my_zipcode.population
# print(prob_american_indian)
prob_asian = my_zipcode.population_by_race[0]['values'][3]['y'] / my_zipcode.population
# print(prob_asian)
prob_hawaiian = my_zipcode.population_by_race[0]['values'][4]['y'] / my_zipcode.population
# print(prob_hawaiian)
prob_other_race = my_zipcode.population_by_race[0]['values'][5]['y'] / my_zipcode.population
# print(prob_other_race)
prob_two_or_more = my_zipcode.population_by_race[0]['values'][6]['y'] / my_zipcode.population
# print(prob_two_or_more)
```

```
        random_number = random.randrange(0, 100000) / 100000
        #print("Random number is:", random_number)
        #print("The sum of the probabilities is:", (prob_hawaiian + prob_black + prob_american_indian + prob_asian + prob_white +
prob_two_or_more + prob_other_race))

        if (random_number < prob_white):
          white_count += 1
        elif (random_number < prob_white + prob_black):
          black_count += 1
        elif (random_number < prob_white + prob_black + prob_american_indian):
          american_indian_count += 1
        elif (random_number < prob_white + prob_black + prob_american_indian + prob_asian):
          asian_count += 1
        elif (random_number < prob_white + prob_black + prob_american_indian + prob_asian + prob_hawaiian):
          hawaiian_count += 1
        elif (random_number < prob_white + prob_black + prob_american_indian + prob_asian + prob_hawaiian + prob_other_race):
          other_race_count += 1
        elif (random_number < prob_white + prob_black + prob_american_indian + prob_asian + prob_hawaiian + prob_other_race +
prob_two_or_more):
          two_or_more_count += 1
        else:
          print("You have fallen through the race test crack.")
          print("Random number is:", random_number)
```

```
    if ((my_list[18][:5])) in heat_map_dictionary:
      heat_map_dictionary[(my_list[18][:5])] += 1
    else:
      heat_map_dictionary[(my_list[18][:5])] = 1

    #print(heat_map_dictionary)



if (my_list[8].startswith('(')) or (my_list[8] == ".") or (my_list[8] == ",") or (my_list[8] == ""):
    # can't analyze these for name
    punch_count += 1
  else:
    # print(my_line)

    # if (my_list[8] == "('Russ')William") ?
    #    my_gender = male_count +=1
    # Removed the ('Russ') from this line
    # but really should write a test for this.

    # wrote a test for "," and for ".", and for "".
```

```
    my_gender = (detector.guess(my_list[8]))
    if my_gender == "male":
        male_count += 1
        # print(my_list[8])
    elif my_gender == "female":
        female_count += 1
        # print(my_list[8])
    elif my_gender == "unknown":
        unknown_count += 1
        unknown_names.write(my_list[8])
        unknown_names.write("\n")
        # print(my_list[8])
        # write these names to a file
        # for further processing



total_count += 1
```

```
        else:
            other_count += 1
            # print("other license type: ", my_list[8])

    print("female:", female_count,
        "male:", male_count,
        "unknown:", unknown_count,
        "punched out:", punch_count)
    print("white:", white_count,
        "black:", black_count,
        "american indian alaskan native:", american_indian_count,
        "asian:", asian_count,
        "native Hawaiian or pacific islander:", hawaiian_count,
        "other:", other_race_count,
        "two or more races:", two_or_more_count,
        "zipcode fail:", zipcode_fail,
        "total race count:", (white_count + black_count + american_indian_count + asian_count + hawaiian_count + other_race_count +
two_or_more_count + zipcode_fail))
    print("total count:", total_count,
        "other license type", other_count)


    #print("Zip code array: ", heat_map_array)
```

```
my_file.close()
unknown_names.close()



heat_map_dataframe = pd.DataFrame.from_dict(heat_map_dictionary, orient='index').reset_index()
heat_map_dataframe.columns = ['ZCTA5CE10', 'licensees'] #column headings for ca_california_zip_codes.geojson
#heat_map_dataframe.columns = ['zip', 'licensees']    #column headings for SanDiego.geojson

#print(heat_map_dataframe)  #check to see if it looks right


#zip_geo="92130.geojson" #worked
#zip_geo="SanDiego.geojson" #worked
zip_geo="ca_california_zip_codes.geojson" #worked
```

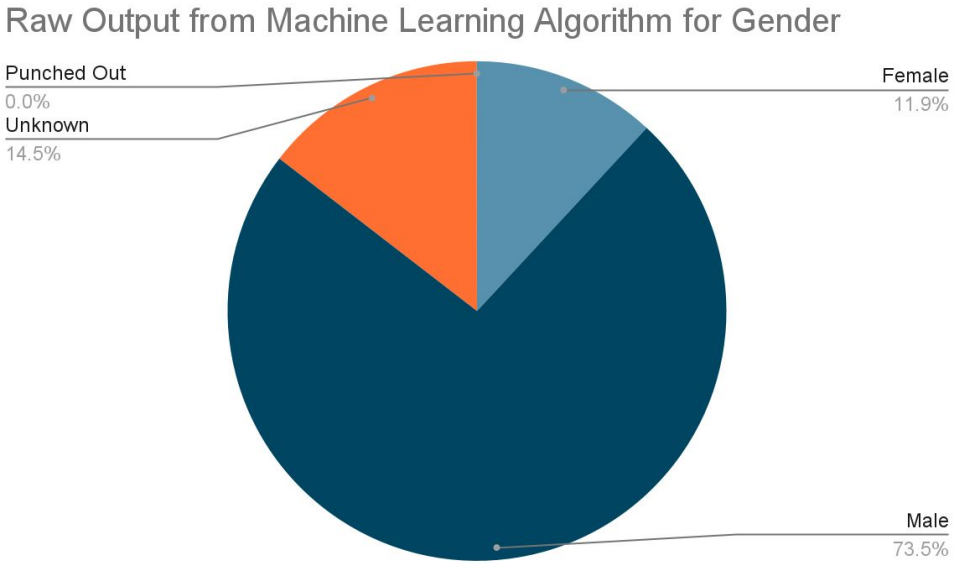
```
m = folium.Map(location=[33, -117], zoom_start=5)


  folium.Choropleth(
    geo_data=zip_geo,
    name='choropleth',
    data=heat_map_dataframe,
    #columns=['zip', 'licensees'],
    #key_on='feature.properties.zip',
    columns=["ZCTA5CE10", "licensees"],
    key_on="feature.properties.ZCTA5CE10",
    fill_color="YlGn",
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name="Heat Map",
  ).add_to(m)

  folium.LayerControl().add_to(m)


  m.save('heatmap.html')
  webbrowser.open('heatmap.html', new=2)
```

The Results

female: 106089
male: 655260
unknown: 129426
punched out: 259

white: 690909
black: 62444
american indian alaskan native: 7584
asian: 39099
native american or pacific islander: 1644
other: 37795
two or more races: 24959
zipcode fail: 26600

total count: 891034
other license type 633475

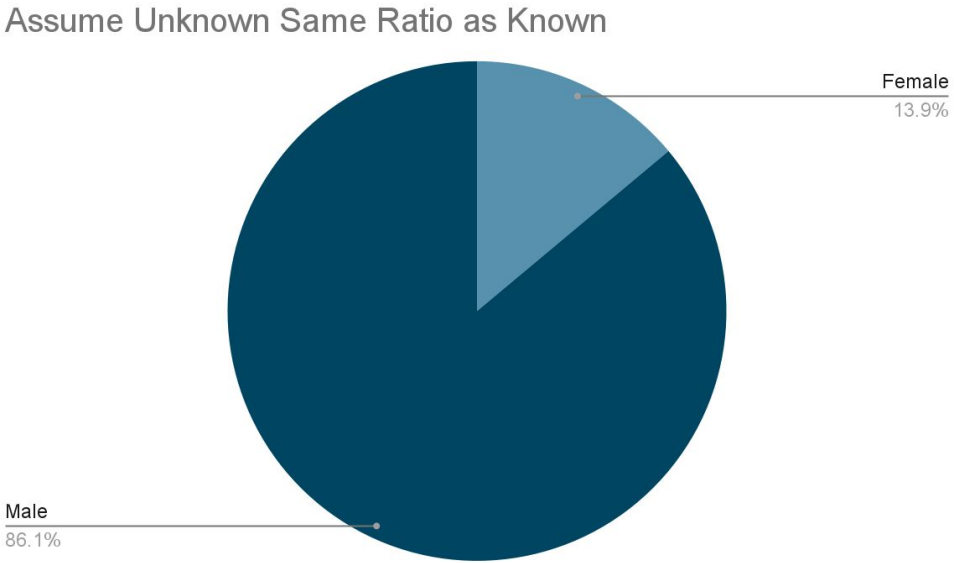


Raw Results

female: 106089
male: 655260
unknown: 129426
punched out: 259

white: 690909
black: 62444
american indian alaskan native: 7584
asian: 39099
native american or pacific islander: 1644
other: 37795
two or more races: 24959
zipcode fail: 26600

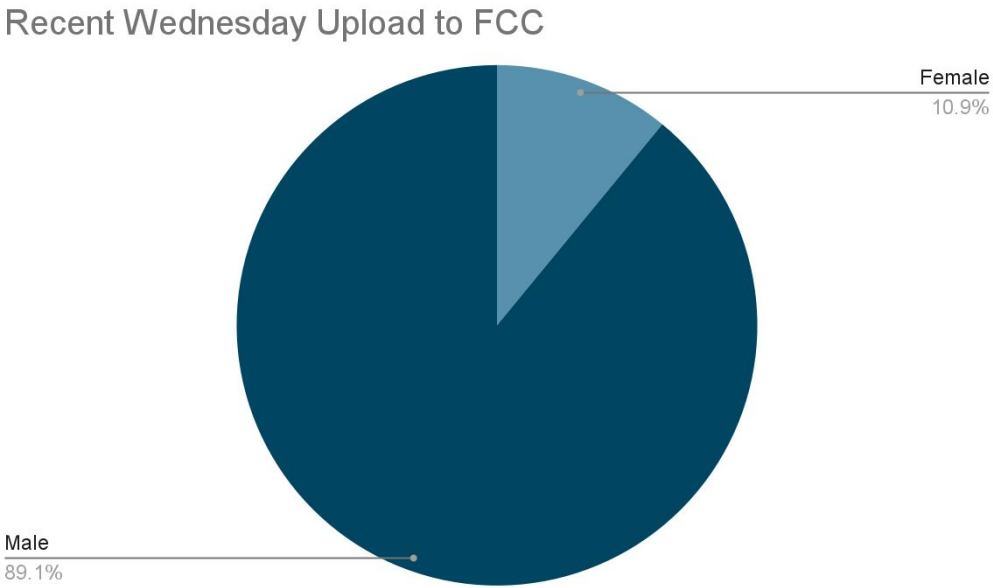
total count: 891034
other license type 633475



Raw Results

No change in gender ratio for active US individual licenses since 2005. A slight decline, from 14.85% to 13.9%.

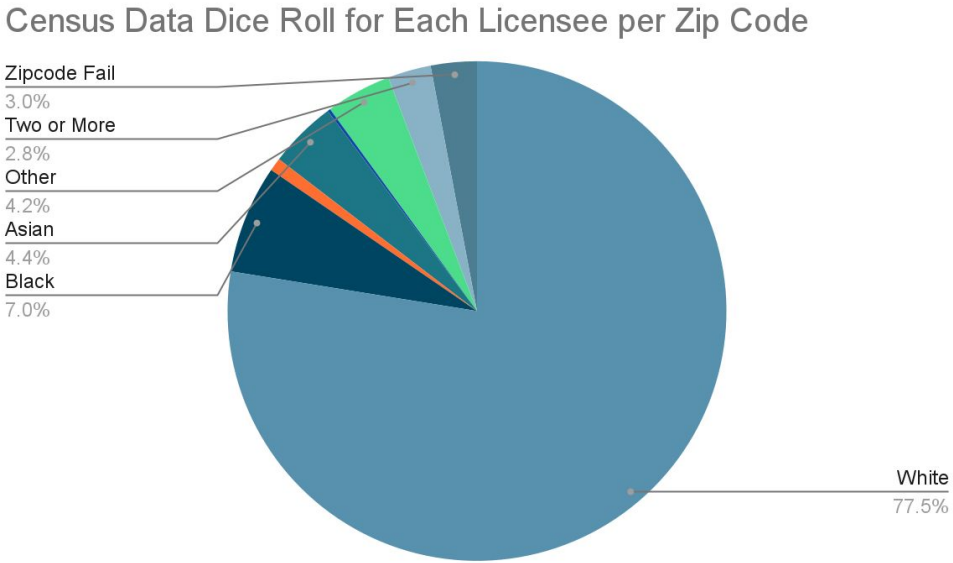
No significant difference in very recent license grants.



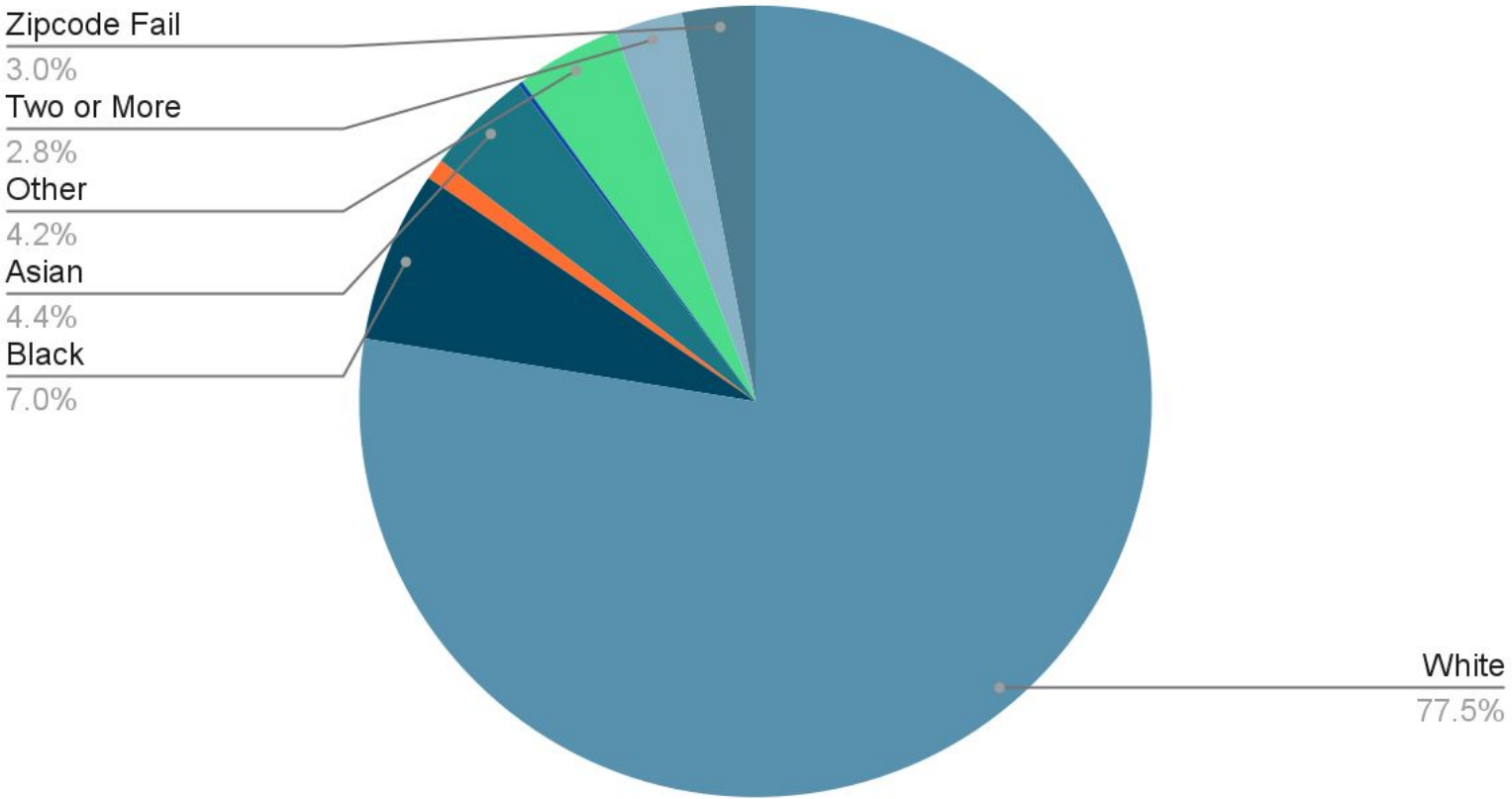
female: 106089
male: 655260
unknown: 129426
punched out: 259

white: 690909
black: 62444
american indian alaskan native: 7584
asian: 39099
native american or pacific islander: 1644
other: 37795
two or more races: 24959
zipcode fail: 26600

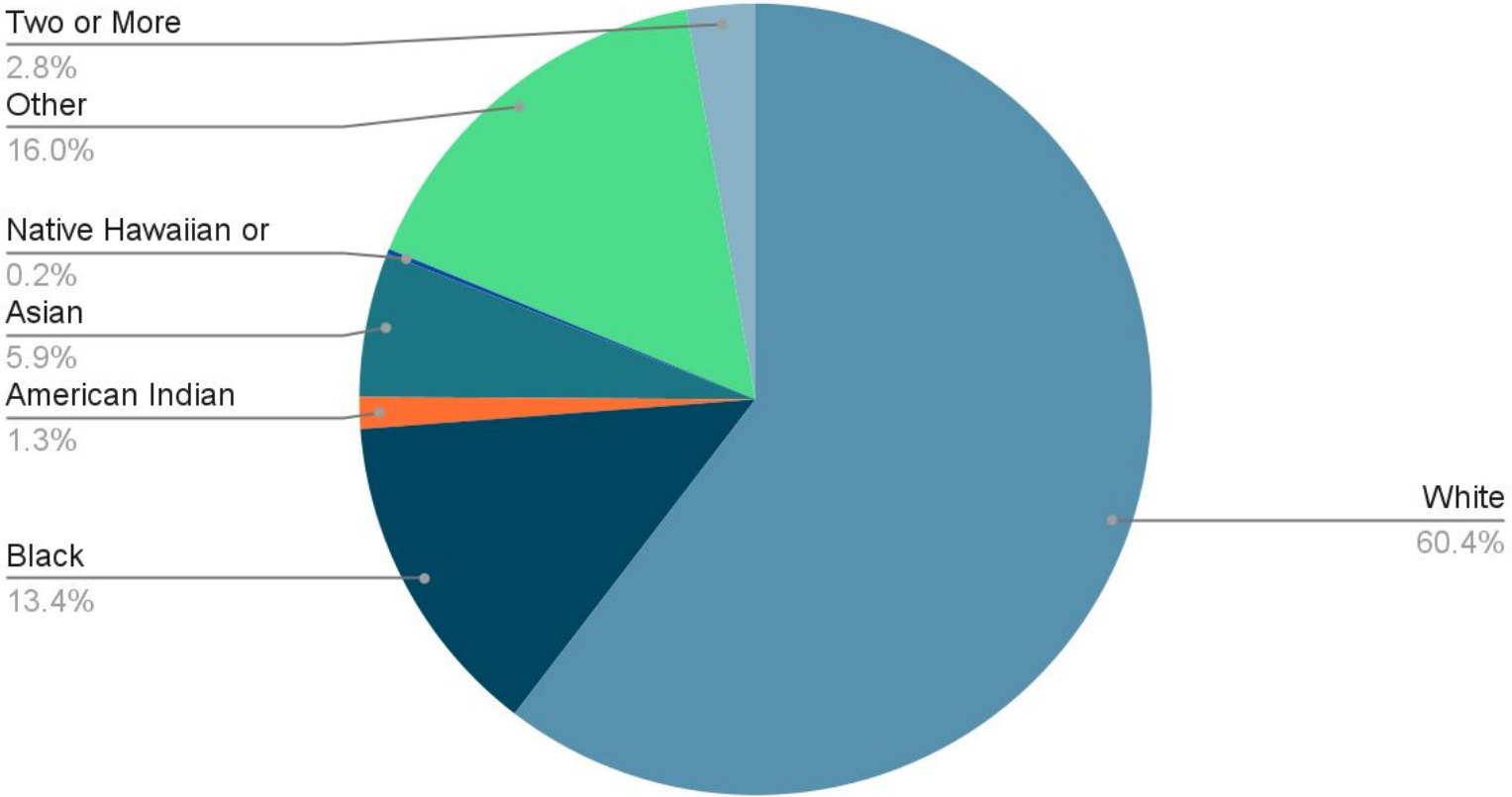
total count: 891034
other license type 633475

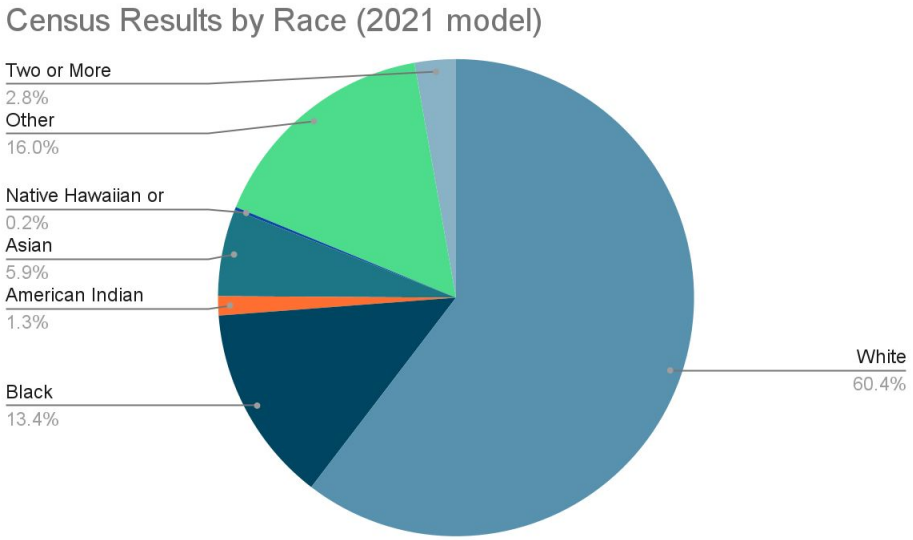


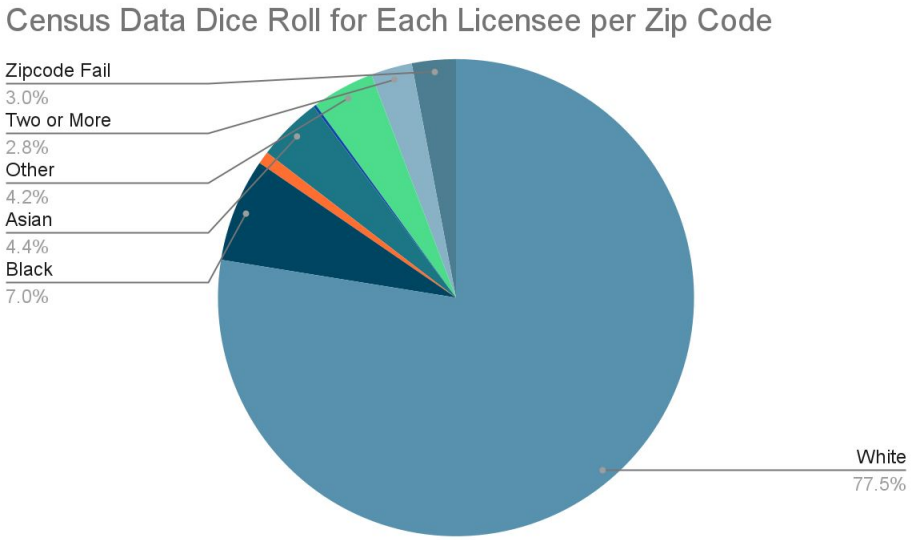
Census Data Dice Roll for Each Licensee per Zip Code



Census Results by Race (2021 model)





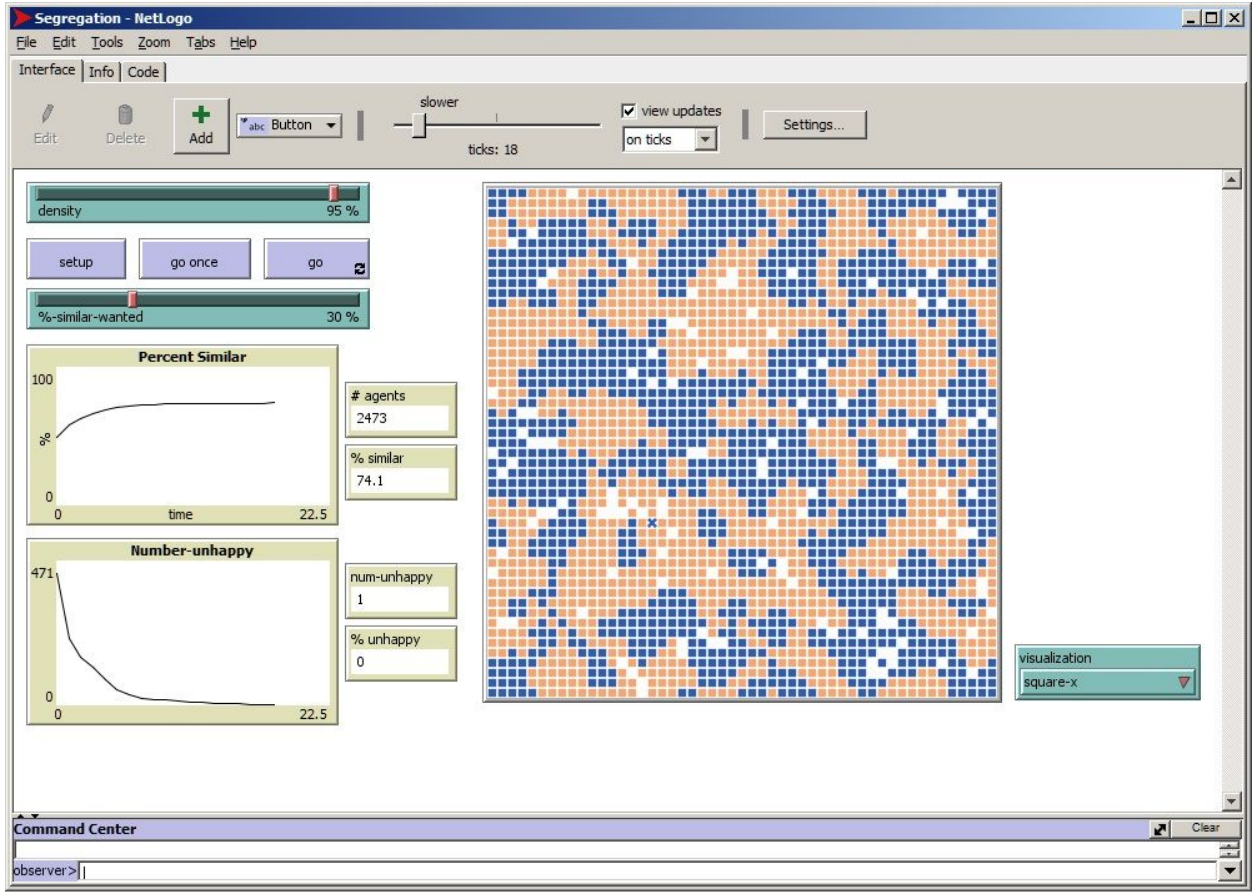


This is the <u>best case model</u>, since it does not take into account **sorting** or **peer effects**.

“Micro motivations do not equal macro behavior.”

Density, percent similar wanted, and the ratio of the categories are all critically important in modeling segregation.

Mild individual preferences lead to significant amounts of segregation. Small categories of people do not survive exclusionary segregation because setting up parallel cities, clubs, associations, societies, non-profits, companies, etc. can easily be out of reach if there simply are not enough people to go around. Peer effect can eliminate participation over time.

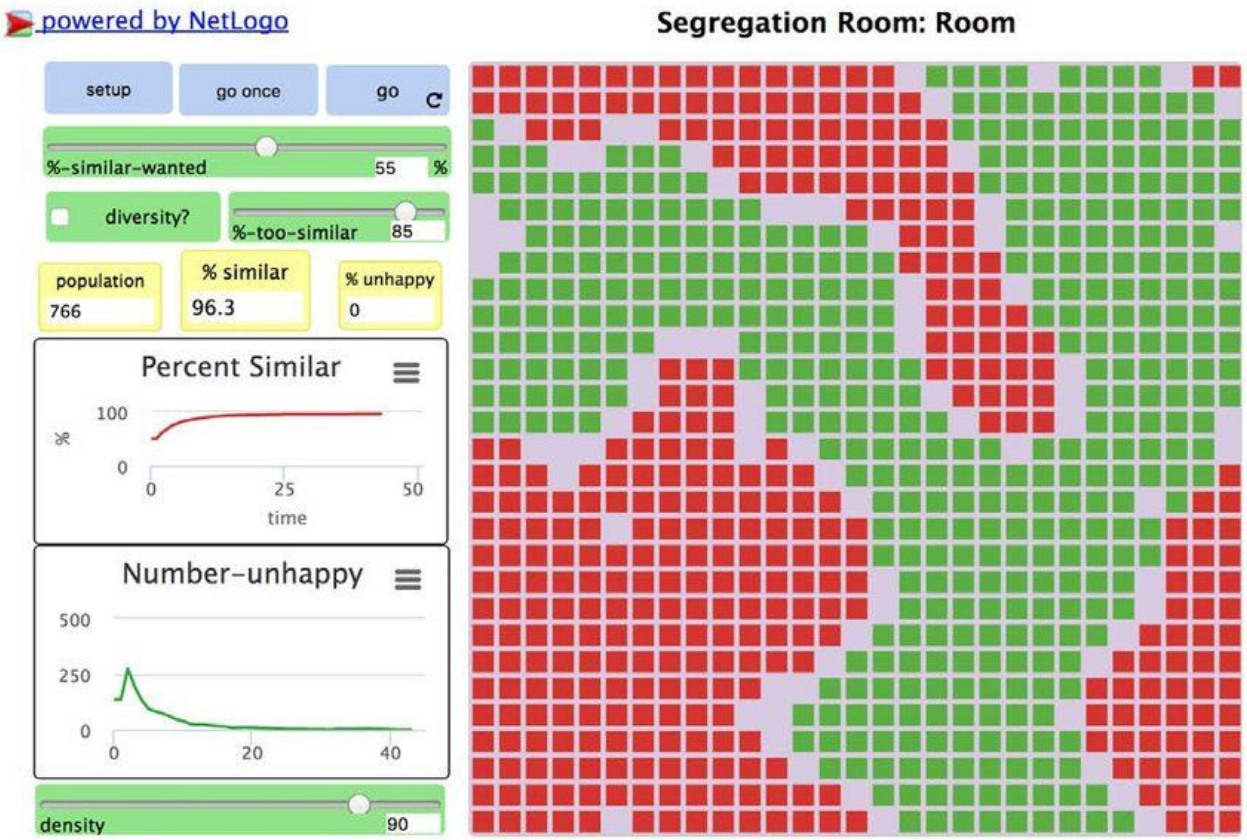


We have very lopsided demographics in traditional radio clubs. More extreme than what the models presented here predict.

There are better demographics in ad-hoc and interdisciplinary radio-related groups.

Amateur radio has an obligation to be of real public and educational service to the general public.

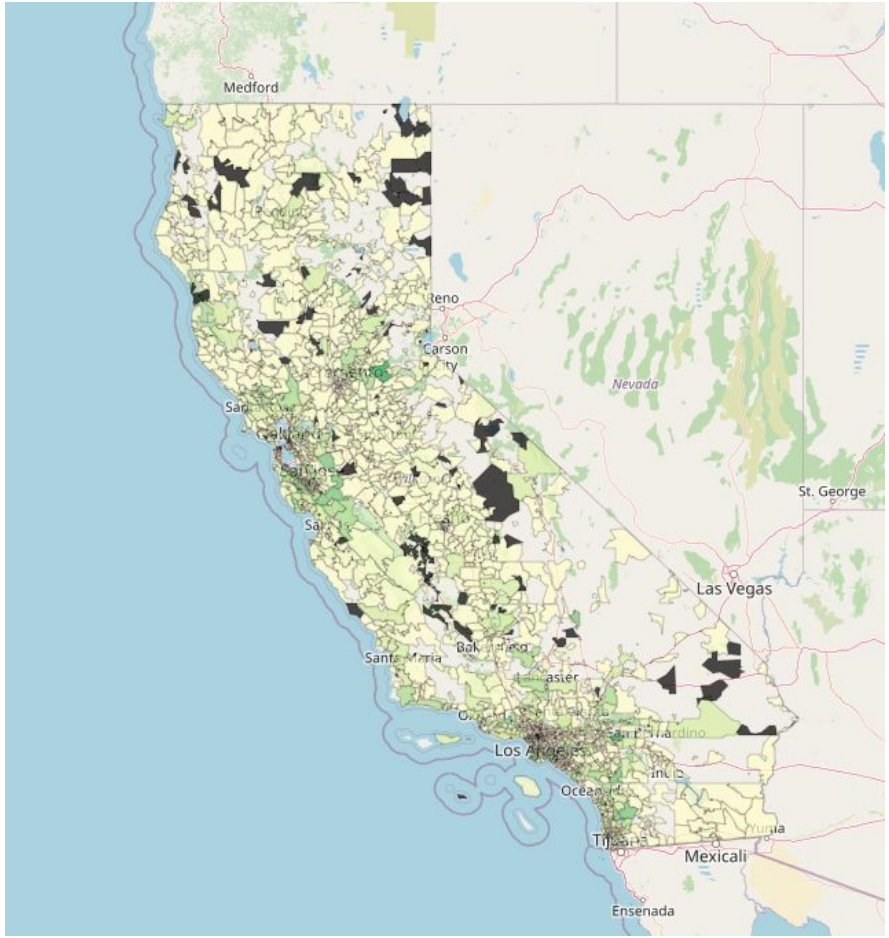
So, what do we do about this?

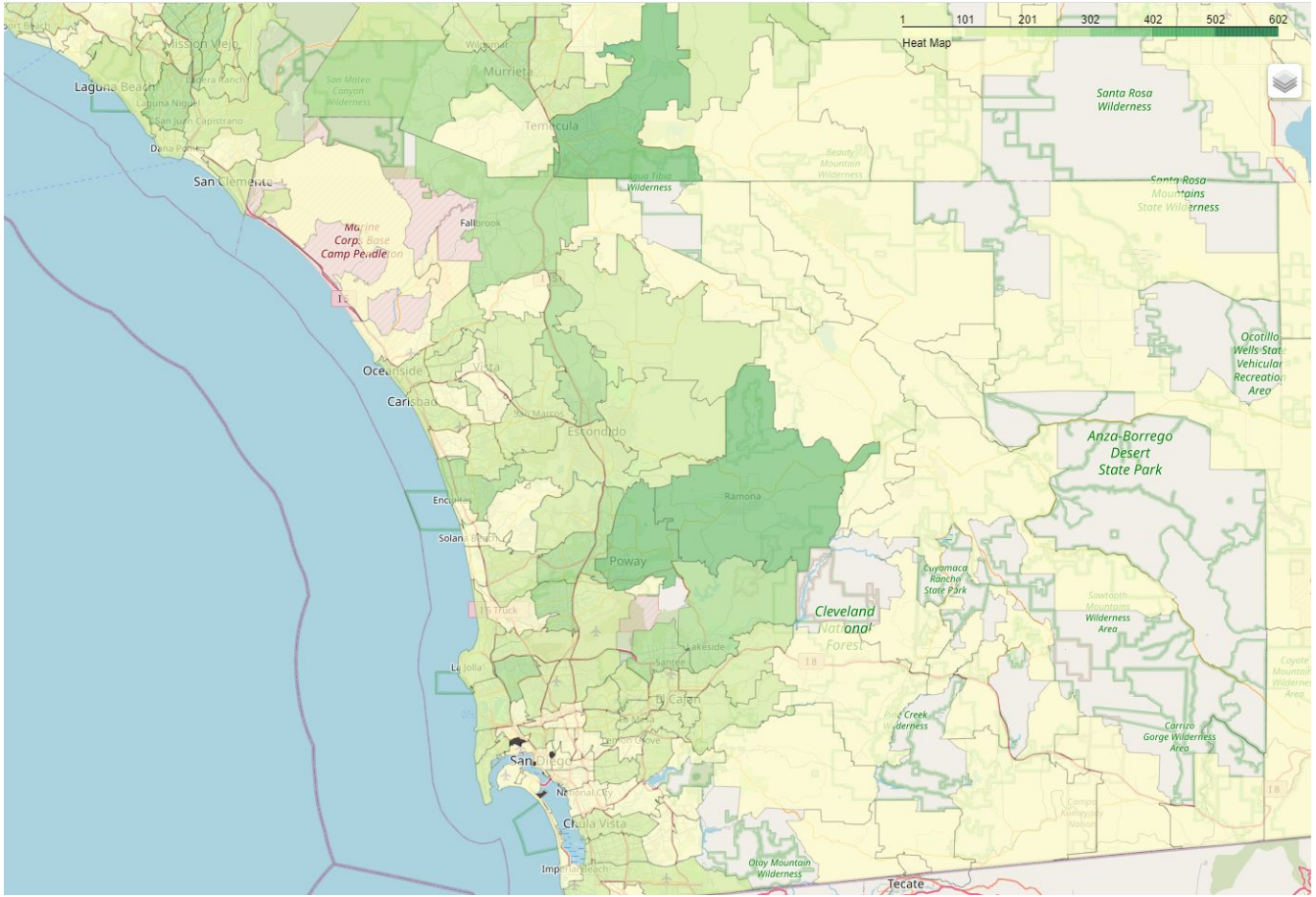


Using the *folium* Python library, we can make choropleth maps that vary color intensity per zip code depending on how many licensees are in that zip code.

We can see concentrations of licensees this way. Here is the first draft, the state of California. You can see some of the disadvantages of working directly with zip codes and not (yet) converting to proper geotagging.

<u>Goal</u>:
Interactive map of the United States.





Where does PARC meet for club meetings?

Carlsbad.

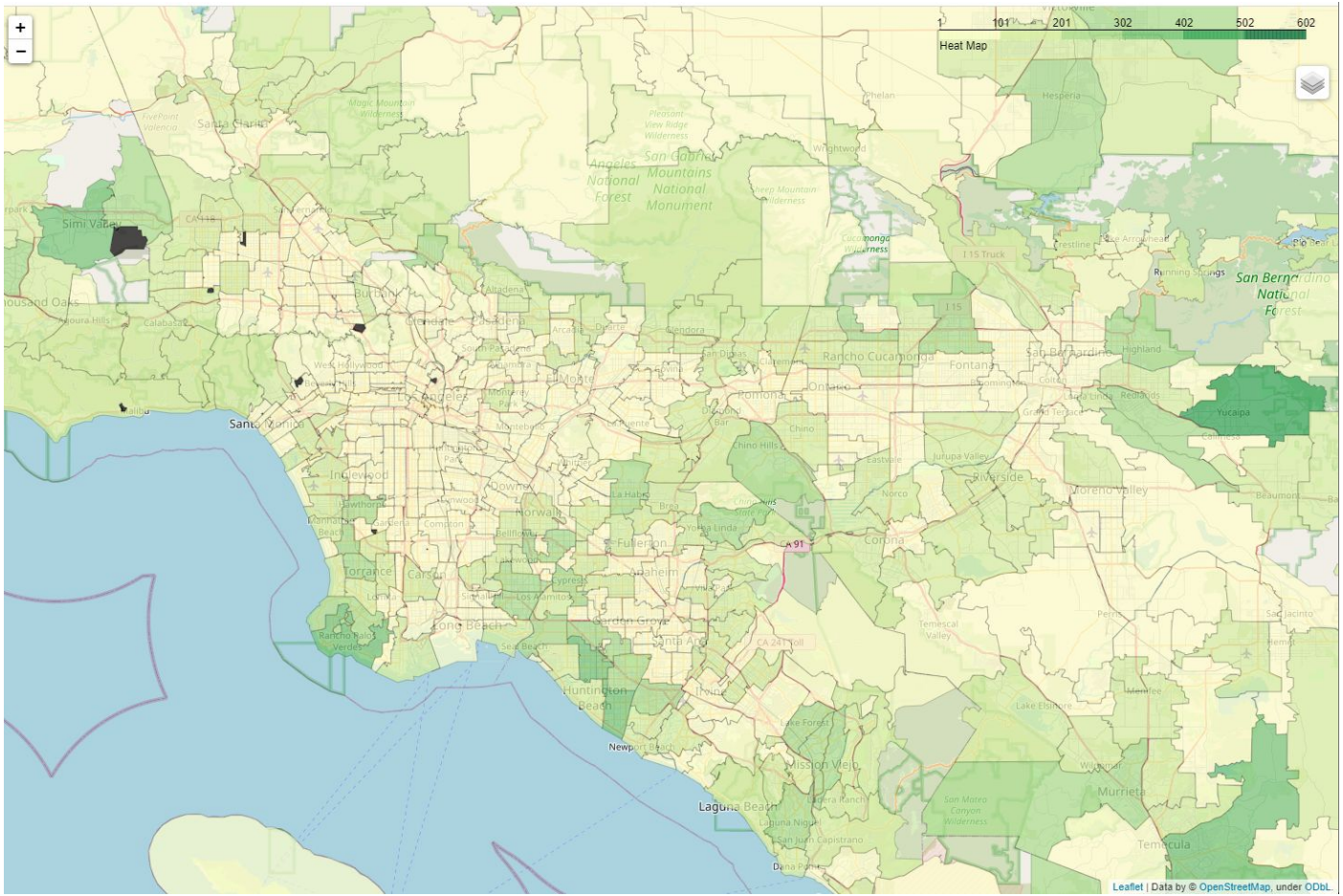
Where are the hams?

Poway, Ramona, Encinitas, Carmel Valley, Clairemont, Mira Mesa, Encinitas, Lakeside, Santee.

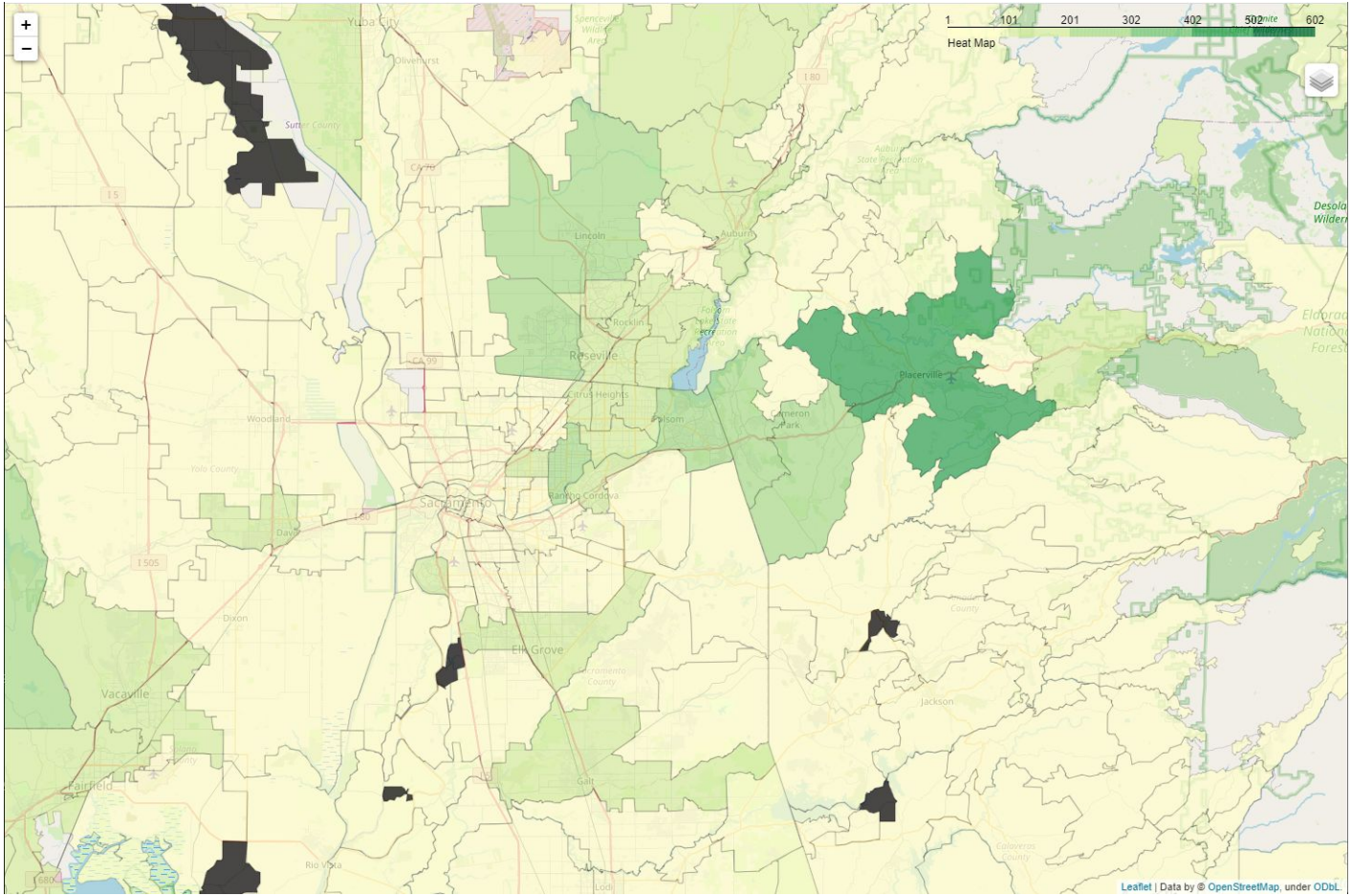
Should the club consider another location for meetings?

What if those areas already have a local club?

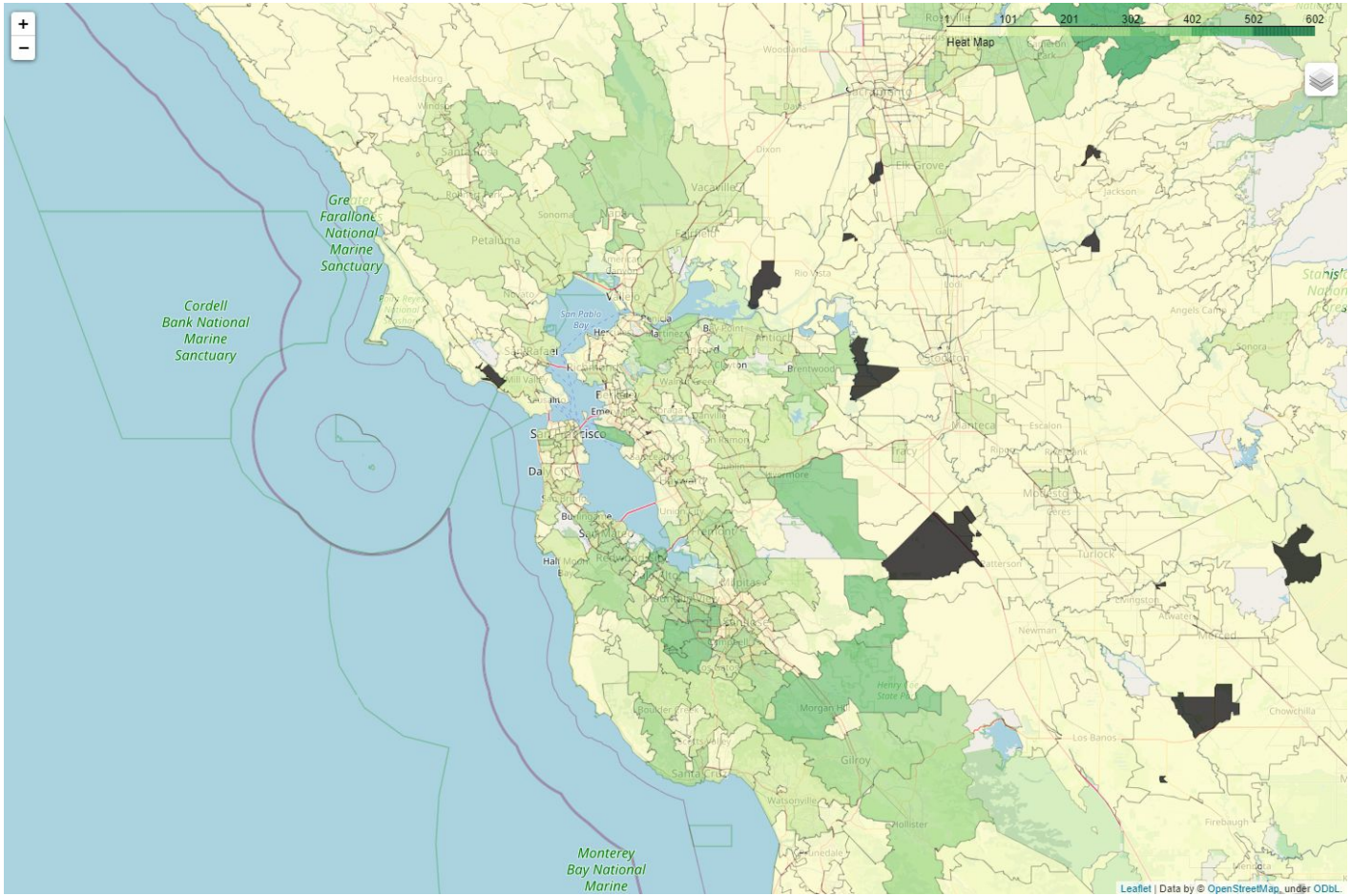
Displaying Results



What is up with Yucaipa?



What's the story on Placerville?



Here's San Francisco.

Displaying Results

Next Steps

What next?

- ➔ Work to get the US FCC to directly provide demographic information. Preferably with more privacy, not less. Can we get demographics without someone like me handling *your* data?
- ➔ Address the shortcomings of using zip codes instead of geotagging.
- ➔ Continue to refine and expand the demographic models to extract more value.
- ➔ Speak up about power structures in amateur radio that are not serving the general public.

Questions?

Repository of Work

https://github.com/Abraxas3d/Demographics

Pull requests, forks, questions, improvements, and contributions are all welcome!

@iamdevloper Jan 8
My coding style would be best described as “there appears to have been a struggle”