Who We Are

Michelle Thompson W5NYV

5379 Carmel Knolls Drive
San Diego, CA 92130

w5nyv@yahoo.com

Who We Are

Michelle Thompson W5NYV
28 December 2022

Introduction

We have a problem in amateur radio. We do not have easily-obtained demographic data and information for US amateur radio licensees.

Peter Drucker is a name that should be familiar to those of you in management. He is responsible for the concept of "You can't manage what you can't measure". He went beyond this very pithy saying in his writings and advised us all that we cannot know whether or not we are successful unless success is defined and it is kept track of. We cannot improve what we cannot, or will not, measure.

What have we done so far with tracking US Ham Demographics?

Well, 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, serve, and enhance the communications skills of the general public, not really knowing "who we are" is problematic from both a regulatory and a practical point of view. How can we continue to justify our spectrum if we cannot show that we are authentically serving the general public?

Are the impressions we have about our demographics 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 Science Technology Engineering and Math (STEM) value, emphasis, and mandates, underrepresentation and exclusion has real economic and national security implications.

Prior Work

There is a good study from 15 March 2005 by Ken Harker, WM5R called "A Study of Amateur Radio Gender Demographics".

http://web.archive.org/web/20070223193600/http://www.arrl.org/news/features/2005/03/15/1/?nc=1

For all amateurs in the FCC database in 2005, he found that 13.60% were female, and 77.94% were male. A number had first names that led to a

classification of Uncertain (8.46%). He adjusted the percentages, and estimated that the Amateur Radio population in the United States at that time was 14.85% female and 85.15% male. This was a slightly higher female percentage than the ARRL surveys indicate, and almost three times as high as the percentage of ARRL members who were women.

Remember these numbers.

Kai Siwiak, KE4PT is the editor of QEX, among other duties, at ARRL. 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

90% are older than 61. Half of that 90% is older than 70. Two thirds were retired. 9% were between 51 and 60 years old.

He makes a very clear call here in this editorial for all hands on deck to address age diversity in QEX subscription base. He acknowledged in his column that the term "Elmer" is problematic, and he proposed “Mentor” as a replacement.

There is a good study from 2021 about demographics in CW contesting. It shows what appears to be a linear mapping of time passing to age distribution in CW and Phone contesting. If you haven't read this article, please do.

By Frank K4FMH
https://www.amateurradio.com/the-secretstorm-approaching-cw-contesting/

The Radio Society of Great Britain has information about the demographics of

people taking license exams in Great Britain. Age and gender data is available. Half of the exam candidates were between 40 and 59 years of age, and 95% were male. The reports are excellent and, like the other prior work discussed in this article, are available online.

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

In prior work, we can see some clear indications of lopsided demographics in amateur radio. But is this necessarily bad? I'm calling it a problem, but are there really any differences experienced by minorities in amateur radio?

Yes, there are big differences in how minorities experience things in amateur radio. For example, I experienced a big difference in the type of mail I received from two co-authored ARRL QEX articles. One was about 3d printing microwave horn antennas, and the other was a recommended strategy for AMSAT-NA to take full advantage of current space industry realities. In both cases, I collaborated with a man that I regularly work with. In both cases, my name and the male co-author's name were listed by each other, and contact information was presented for both without any significant differences in the way a reader would see it.

In both cases, the mail received by the male author was respectful, positive, and complimentary. In both cases, the mail received by me was aggressive, dismissive, and interrogative. There was at least one extremely harsh message with personal insults and expletives. I "should never spoil the magazine with" my "dreck ever again". In some cases, the mail received by my male co-authors bordered on fawning.

I attended a meeting where one of my co-authors was often present. When I brought these emails to the meeting and read them out loud, my co-author was astounded. He had no idea why anyone would respond to the article this way,

and he had not received anything like the messages. In fact, in every other case where he'd written an article, for publications like the proceedings of the Microwave Update, he'd never received anything negative, personally offensive, or unfriendly.

It was eye-opening to me that neither of my co-authors in amateur radio had ever even considered contacting law enforcement because of messages they received after publishing technical articles. They have never received any violent threats for speaking or writing about amateur radio. I have received violent threats after publishing technical and regulatory work.

In discussions with amateur operators over the years, I've been assured that women "just aren't interested in tech", and that women "should be seen and not heard". I have been ordered "silenced" by men who run amateur radio funds, was told that a woman "can't be above a man" in a volunteer org chart at AMSAT because of what it says in the Bible, that white people are more focused and organized and ham radio needs focus and organization, that my efforts to achieve some racial integration in amateur radio are "annoying", and that "amateur radio just doesn't suffer any harm from being homogenous".

Elected amateur radio leaders have told me that technical volunteer teams composed of all white men from very similar backgrounds can "simply work faster and better together", and that including people that look and think differently, like me, would just slow things down. I have been told told "everything was fine until you showed up", as if pointing out exclusionary and unprofessional behavior in amateur radio organizations was a worse offense than the behavior itself. I have been publicly retaliated against for privately complaining about clearly sexist language and decisions.

Sexism, racism, and bigotry exist in amateur radio. These problems are real and they come from both leadership and the rank and file. These problems cause harm. It's natural to wonder if the problems and the demographics are related.

Could we revitalize amateur radio with diversity?

Is diversity the answer? Are diverse communities in a better place than homogenous ones? Lucky for us, there are a number of studies that show the advantages of healthy diversity and inclusion.

A good survey article from Harvard Business Review is
https://hbr.org/2016/11/why-diverse-teams-are-smarter

From that article, we see that diverse teams are much more likely to evaluate opportunities correctly. Diverse teams give results above industry mean. This is a very nice "free" return on investment.

Something close to my heart, since my primary focus is research and development for open source amateur radio, is 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."

This finding is from 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

When we talk about demographics, we are asking questions about bigotry, racism, and sexism. There are many assumptions about these negative behaviors. Some of those assumptions are incorrect. For example, individual attitudes do not equate to large-scale demographic features. What does happen is that large-scale demographics can create environments where bigotry, racism, and sexism can seem normal, reasonable, and make sense. Individual preferences may not

create an organization or community that is, for example, ageist. But, over time, members of that organization or community can behave in ways that exclude young people, even when attracting young people is a stated priority. Members of the organization or community may say that they "never see" examples of exclusionary behavior, and that they have "no idea what all the fuss is about". When there are literally no young people, no non-whites, or no women, then the members may indeed never see what made those members leave and what keeps them from coming back.

So what is the solution for a lack of demographic data? We have some prior work. What can we add?

With basic modeling, demographic data about US amateur radio licensees can be constructed. Gender can be determined 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. This work was carried out.

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. The University of Michigan has the most public-facing presentation of this scientific technique if you would like to learn more about it.

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. For example, the way the CW contesting work cited as prior work was constructed can be extended to other well-defined affinity groups within amateur radio.

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, and 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.

A Walk-Through of the Computer Code

Python scripts written for this article, and other information about the study, can be found at https://github.com/Abraxas3d/Demographics

With rare exceptions, when we sit down to solve any sort of problem, we don't do everything from scratch. When sewing clothes at home, we often use patterns and instructions written by experienced drapers and we use fabric designed and manufactured in advance. We don't make our own sewing machines from scratch or forge our own needles. When we do something like building a radio kit, we buy parts made in a factory and we follow a schematic designed by someone else. We quite often have to troubleshoot problems along the way, and many of us find it fun to modify or kit bash. When we work on a car, we use tools that we bought, maybe from Sears back in the day, and we follow along the Haynes manual, checking off the steps.

In the first section of the code we bring in some very useful hand-tools. In Python, the language used for this computer program, we gather up a list pre-made blocks of code we're going to use. To do this, we use "import" statements. The imported code lets us move around in the directory structure on a computer

and open files. It lets us generate some random numbers. We're going to present the data in a web browser, and we're going to make some zip code maps. We import tools to accomplish those things. We use something called Pandas which gives us a useful structure called a data frame. This structure bolts up to a library called Folium. They work together. We import a machine learning tool for gender detection and we import a zip code search engine that returns census data. Zip code in, census data out.

We then create a set called active_licenses. A set is a collection in Python where the order doesn't matter, you won't need to edit the item after you add it, and there is no index. It's one of several built-in types in Python. There's nothing in this set when we start out. It's empty.

The code opens up files that we download directly from the FCC. This is from the downloadable databases page.

https://www.fcc.gov/uls/transactions/daily-weekly

One zip file has several files in it, and the one we start out with is the HD, or the header file.

Here is an example of what the lines of data in this file look like.

```
   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||||||||
```

The code attempts to open this file. If it's not there, there's an error. Once the file is open, it picks up the first line. Think of this as opening up a big binder of

this file printed out, and looking at each line. Each line is an entry of data about one US license, or call sign.

We split up the line we're looking at on the vertical bars. This turns the monolithic line into a list. We then take that list and go to the sixth field. Or, the fifth field if you are counting from zero. That's where the 5 in square brackets comes in. Notice that field consists of a single letter. If that letter is A then the call sign on that line is active. It can be expired or canceled, but we are looking for active licenses. We then move one field over, see the square bracket 4 element of our list? and we put that value into the set that we made called active_licenses.

If it's an active license, stash that call sign in a set.

The octothorps or hash sign indicates a comment. In testing the code, I would print out the set of active licenses to make sure it was working.

After we've gone through the entire HD.dat file, we have a set of active call signs.

Next, we set up our gender detector. We are selecting the version trained on US names.

We also set up our zip code database. This requires a download of US census data that is designed to work with this imported module. It should download this database the first time the code is run.

We create a file called unknown_names, to catch the list of names that the categorizer could not figure out.

Next, we set up a scratch pad for making our calculations and counts.

We then try and open the Entity, or EN.dat, file.

This has full address information for all amateur radio entities.

You can see what this file format looks like below.

```
   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||||||
```

Now that we have a list of active call signs, we can go through the EN.dat file and pick out addresses of active US licensees.

If we can open this file, then we go through line by line. For each line, we separate the fields by busting up the monolithic line on the vertical bars. If the call sign in this line is one we determined to be active, and if it's an individual license, then we go on.

The first thing we do, if the line corresponds to an active individual license, is take the zip code from that line and look it up in our zip code census database.

If this lookup fails to have data about population by race, we increment the "zip code fail" variable, and skip ahead. Not every one of the 42,000 zip codes in the database has census data.

If it does have population by race data, then we take the race ratios and make a probability table. If the zip code is 62% white, then 0.62 is calculated as the probability of being white. If it's 7% Native Alaskan, then it's 0.07. We use the census categories for making this table.

All probabilities for a single zip code better add to one. That's a rule of probability theory and it's how we found a bug in the code early in the process,

where we double-counted a category and left another category out.

We generate a random number between 0 and 1 with fine enough resolution to be below any quantization error.

We look for which category that number falls into with some comparisons. We increment whatever category we ended up in with our electronic dice roll. Note that the results may be slightly different every time, because we are generating random numbers and using a simple model of what race this licensee probably is, based on the statistics of your zip code.

Now, we want to make a map of where the hams are. We want to be efficient. We could list all the zip codes and then put tick marks by the zip code of the active ham we found in the EN.dat file. But, we're going to do better than that.

We set up something called a dictionary. This is a Python structure. It has a key and a value. Like a dictionary, you have the word, which is the key, and then the definition, which is the value. You look things up by a key, and usually return the value as the information fetched.

For this dictionary, We have zip code as key, and the number of hams in that zip code as the value.

If we have a zip code we need to put a tally mark by, we first check to see if it's already in the dictionary. If it is, then we increment the value by one. If that zip code was not already in the dictionary, if it's a new one, then we set the value to 1.

This means we have a dictionary with no gaps. The performance of this data structure is pretty fast. Notice it's not stored to a file or anything like that. We're building it as we go in memory.

So, we move on to names and we make a guess about gender. We're still working with the same line from the EN.dat file, but now we pick up the name field. We have some quirks in this database that we are going to punch out, or disregard. If that record has no name, a comma as the name, or a period as the name, or something else that isn't a recognizable name. It's very unusual these days to get a license application without a name, but there are active individual licenses with non-word names. Therefore, we skip those.

We take that name and run it through our gender detector. This is a model created with machine learning, and trained on US names. It assigns either male or female or unknown. Names like Pat, Leslie, and so on fall into the unknown category.

We write the unknown names to a file to make it easier to run more tests on them.

We update our counts for gender.

Then, we print out all the counts.

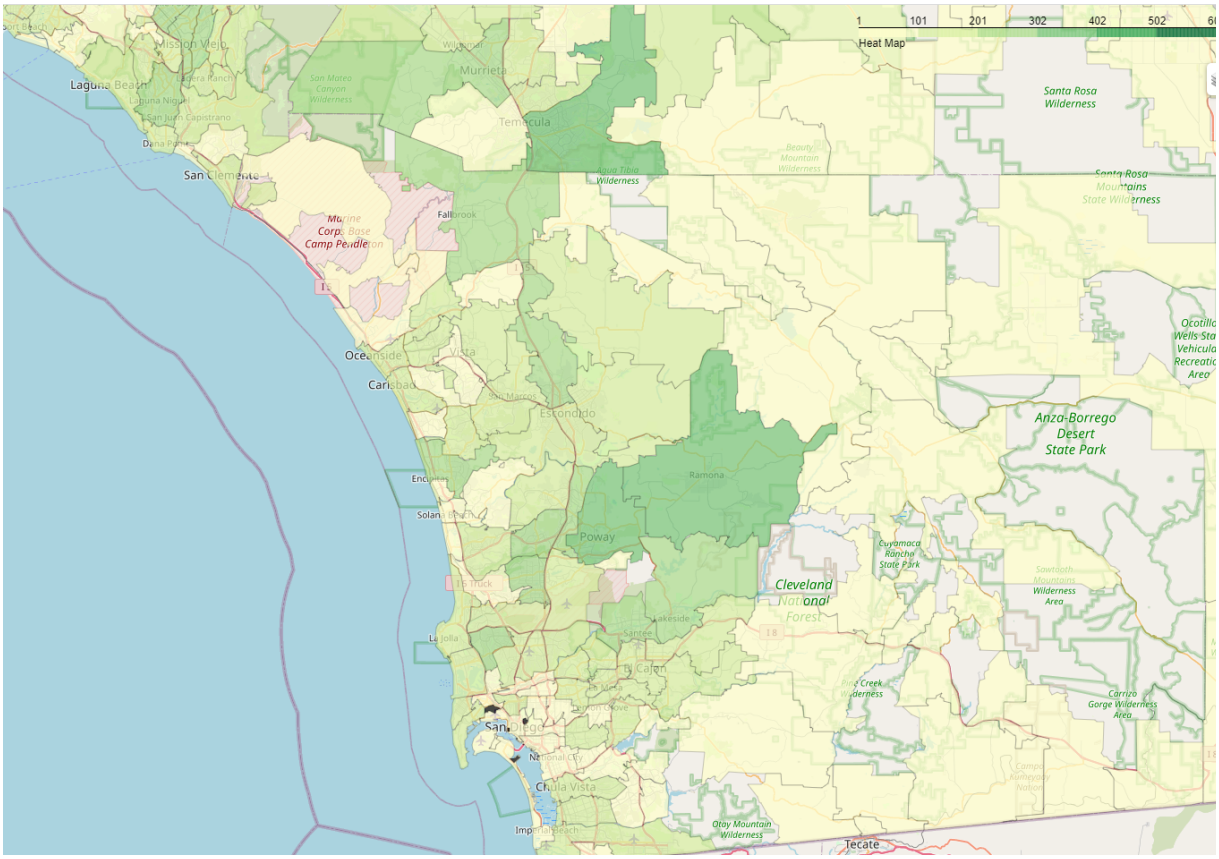
If the license was not an individual active license, which was our first real test, then we skip over all the tests and simply put a tally more in "other license type".

Next, we make a heat map or choropleth. We have a dictionary of zip codes and tally marks for how many hams per zip code. We convert that dictionary into something called a data frame, which can then bolt right up to the maps in the tool we imported at the beginning, called folium. We make sure we put in some headings for our dictionary/dataframe thing we have made. The name of the column in our data must match the name of the column in the map we want folium to use.

We have lots of choices out there for map files. The default example in the

repository for the computer program is the US State of California.

A choropleth is a heat map. It uses color value, light to dark, to indicate the number of items in a region. We set up the folium map, tell it where to center and how far out to zoom, and then tell it what colors to use and where our tally mark file can be found. We save this map as a webpage called heatmap.html, and we open it in a web browser. Here is an example result.



SanDiegoChoropleth.jpg

Results

The Python script written for this article, and other information about the study, can be found at https://github.com/Abraxas3d/Demographics

The program analyzes the information in the downloadable amateur radio databases obtained from the US FCC's website. Gender and race for active individual US licensees is determined.

First, we consider the raw output from the gender detector for the entire database. Female and male names were identified and counted. However, 14.5% were unknown. Names such as "Pat", "Leslie", and "Rusty" belong to a category where they could be either a male or a female name. The Python script wrote these names to a file and we examined them more closely.

After looking at a random sample of those names, and doing our best to track down the person online or on social media, the ratio of the unknown gender turned out to be the same as the ones the detector was more sure of. Assuming same ratio as known, we get 13.9% female and 86.1% male.

Assume Unknown Same Radio as Known.png

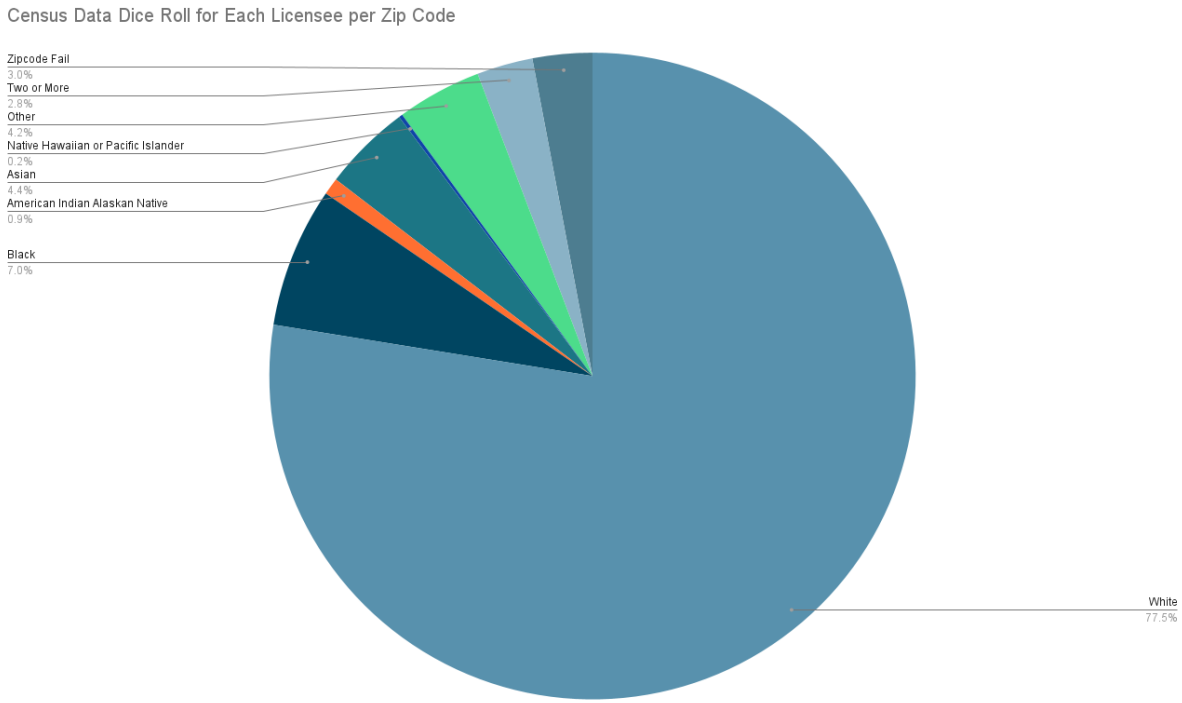
This means there has been little change in gender ratio for active US individual licensees since 2005. There is a slight decline in women US licensees since 2005.

What does that decline look like? Here's the trend over the past 10 years. For this data set, the entities were separated by grant date.

Estimate of Percent Women US Amateur Radio Licensees by Year of Grant Date.jpg

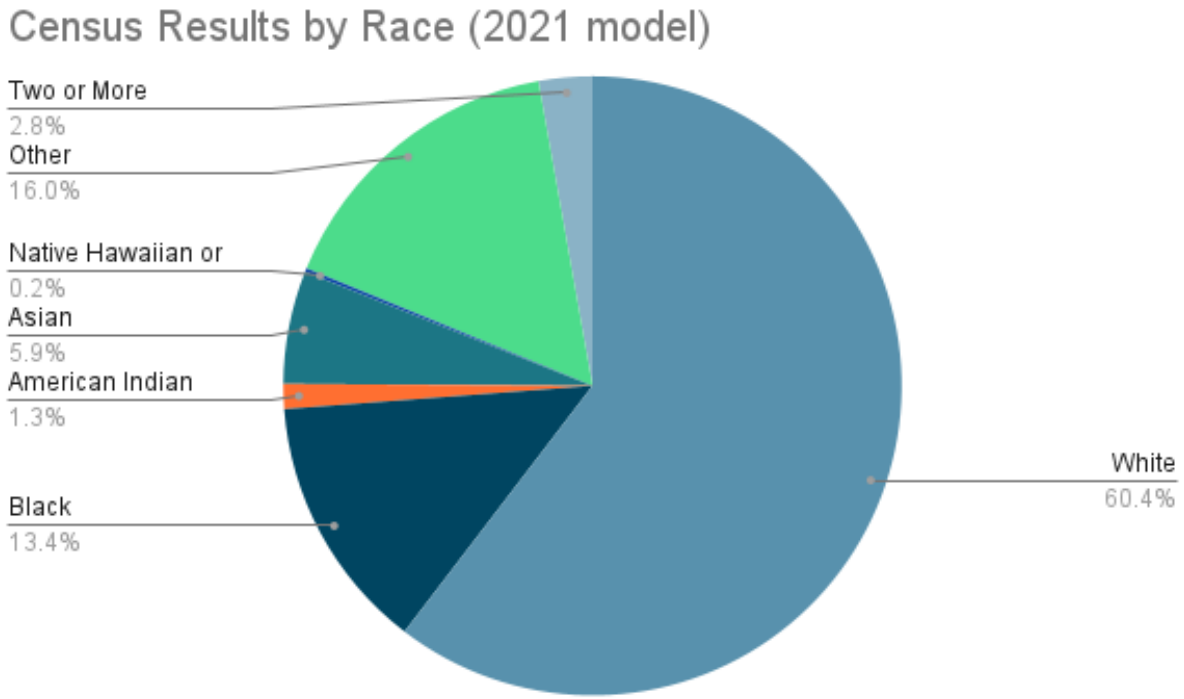
Grant Year	Estimate Women Licensees
2022	12.5
2021	12.5
2020	11.6
2019	13.4
2018	13.5
2017	13.5
2016	13.3
2015	13.6
2014	14.2
2013	14.9
2012	15.3

Next, we consider race. We take the zip code of each active individual licensee, and make a probabilistic guess about their race based on the racial statistics of the zip code they live in.



Census Data Dice Roll For Each Licensee Per Zip Code.png

Here's the census results by race using the 2021 for the entire United States.



Census Results by Race (2021 model).png

Note the zip code fail category, that “two or more” is the same, white has an increase, and all other categories are less.

Does this match what we see in our organizations and activities?

This is a "best case" set of racial results based on which zip codes the licensee lived in. It does not take into account sorting or peer effect.

What happens if we do take into account sorting and peer effect?

First, let's consider sorting. The Schelling model from social science is used to study the effects of segregation. It does not take a very strong individual preference for wanting to live near people that look like you to create segregated neighborhoods. This is a well-studied effect. It does not take a very strong preference for walking into a ham club meeting, seeing no one there that looks

like you, and then leaving and never coming back in order to create a very segregated amateur radio club. People may assume that clubs and organizations in certain hobbies are going to be homogenous, and won't attend because "I already know I'll be unwelcome and I don't need that in my life."

If you use the Schelling model, and you apply it to demographics where populations are lopsided, then the sorting effect means the smaller categories predictably segregate with increased velocity into much smaller spaces, clubs, and organizations when compared to the majority. In some cases, if the demographics are severely lopsided, those "alternate" organizations won't form in the first place. There are simply not enough people in proximity to each other to form functional alternative organizations.

Sorting is the individual choosing to associate with people that look like them. Peer effect is an individual starting to act and behave like the people they are already around. Think "peer pressure" or "tradition" or "well we've always done it that way".

Segregation is well studied in social science. Once you add in sorting and peer effect, then minority members tend to disappear or drop out from active participation at a much higher rate than majority members. Whether minority members are leaving because they were made to feel unwanted, harassed because they were "uppity", or just because they didn't feel comfortable being around people that don't look like they do, the results match what most of us observe in amateur radio. It's almost all men and they are almost all white - especially and particularly in the traditional amateur radio club scene.

So, what can we do about this?

There are ad-hoc and interdisciplinary radio and radio-related groups in the US with visibly diverse demographics with respect to age, gender, and race.

What are they doing right? What can be learned from these groups?

The most commonly noted difference is the a willingness to be firmly proactive about inclusion. The second most commonly noted difference is that diverse communities generally offer clear evidence of being able to carry out repercussions for bad behavior. Inclusion fails if people are recruited and welcomed only to experience unaddressed harassment or exclusion within the community or organization. Instead of looking the other way, suppressing complaints when a person is bullied or excluded, and shooting the messengers, healthy communities have a vocabulary for dealing with discrimination, elevate and discuss complaints, look for root causes, and take action when necessary.

In order to be inclusive, we must be prepared to look at things that we might be taking for granted. For example, we might think reciting the pledge of allegiance at the beginning of ham club meetings is "not a problem." Reciting the pledge is controversial and has been the subject of numerous court cases. People that view reciting the pledge as a turn-off will sort themselves right out of ham radio well before an amateur radio club gets any chance to make reciting the pledge a positive experience through peer pressure. This may be an uncomfortable discussion for an amateur radio club. Not reciting the pledge may be interpreted as disrespectful, unpatriotic, or anti-military. Dropping the pledge won't instantly attract new members.

Another example is the effect of different attitudes by race towards law enforcement. Non-whites in the United States have a much more negative view of law enforcement than do whites. An amateur radio club that has a high degree of activity and interaction with emergency communications, and therefore local law enforcement, will be affected by these views. People may choose to avoid amateur radio clubs that have a strong association with public safety or where law enforcement is uncritically praised.

https://www.pewresearch.org/fact-tank/2020/06/03/10-things-we-know-

about-race-and-policing-in-the-u-s/

Is the solution for amateur radio clubs to stop working with law enforcement? No, it is not. The solution may be for the club to take up the challenge of working to find, serve, and include people that view law enforcement and emergency services differently.

Who are the most overlooked in public service and emergency communications? Who are the most vulnerable in a community? What can amateurs do to build connections with people that do not have regular or positive contact with those doing the emergency communications planning? What groups are attempting to reach vulnerable or underserved communities? Do those groups need communications support?

Are women excluded by default? Are women only welcome if they are passive, agreeable, unambitious, and subservient? If females are welcome in an amateur radio club only when they are children or only if they are "actively mentored", then that is discriminatory. Language and communication must be professional and inclusive, from the club newsletter to the directions given at Field Day. Assumptions that women aren't technical, can't lead, shouldn't be taken seriously, must be "ladylike", or are there primarily to entertain, amuse, and serve men are harmful.

A final example is youth outreach. Attracting youth is quite often a stated goal of amateur radio clubs. What is one of the most common ways we do this? We invite Scouts. There's absolutely nothing wrong with inviting Scouts to meetings and doing things like supporting the Boy Scout Radio Merit Badge. However, if this is the only way that youth are invited to participate in amateur radio clubs, then that club's youth outreach is limited to approximately 2% of school-age children. Boy (and Girl) Scouts in the US are also overwhelmingly white.

The point is that inclusion requires work, not wishes. Feeling vaguely uneasy

about a (practically) all-white, all-male, all-elderly radio service is our starting point, not our end game. Given the basis and purpose for the amateur radio service listed in Part 97, actively addressing the lopsided amateur radio licensee demographics is a necessary thing in order for amateur radio to be of benefit to the general public.

The software that produced the demographic results is from an open source project called Who We Are. Open source means you are able to recreate everything I have done, based on the published code. Open source contributions in order to improve and expand the project are welcome.

You can reach me at w5nyv@arrl.net