Untitled

# Package to Process FCC Fixed Broadband Datasets

## Installation

You can install the development version of processFCC like so:

install.packages("devtools")  
library(devtools)  
install\_github("kdmulligan/processFCC")  
library(processFCC)

## Background of FCC Fixed Broadband Datasets

The goal of processFCC is to process the fixed broadband data sets from the [Federal Communications Commision (FCC)](https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477). The FCC is the federal agency responsible for implementing and enforcing America’s communications laws and regulations. They regulate interstate and international communications by radio, television, wire, satellite, cable, and internet in all 50 states, D.C., and U.S. territories.

The FCC Broadband Data is from facilities-based broadband providers who are required to submit to the FCC biannually for any locations where they offer internet service at speeds greater than or equal to 200 kbps. Prior to 2022, this data was collected using Form 477 where information about where and what type of Internet access is provided at the census block level. However, this form was discontinued in in December 2021. Instead, broadband providers now submit information about their broadband internet access services in the broadband data collection system. Within this system, data is reported more granularity than the census block level. Although a census block level version of this data set is still available in the new format. For more information about the filing system, please reference the form resources [website](https://www.fcc.gov/economics-analytics/industry-analysis-division/form-477-resources).

Because of these changes in reporting, one of the first things you will need to do is decide which time points of the FCC fixed broadband data you wish to use and process. Dates from December 2021 and prior use the “old” format whereas dates June 2022 and beyond use the “new” format. For both formats, fixed broadband data is available at two time points per year: June and December. Use new\_or\_old\_FCC() to check which dates are available and also the format (new or old) that should be used to process the data. The output of the function provides the available dates, the format, the function to use to process the data, and the website if you would like more information.

The FCC Broadband Data comes from Form 477, which facilities-based broadband providers are required to submit to the FCC biannually for any locations where they offer internet service at speeds greater than or equal to 200 kbps. Fixed providers report Census Blocks where they provide service, along with information about technology type and speeds. Thus, fixed provider data is on the Census Block level, which is the smallest unit of census geography. A column in the data indicates whether the broadband was provided to a residential or business location–the processFCC package only considers residential broadband. Importantly for interpretation, if a provider indicates they provide broadband to a Census Block on FCC Form 477 it does not mean every location in the Census Block has access to the broadband service. It simply means at least one location in the Census Block has access to broadband at the reported speed/technology. It is important to consider this as it may lead to over-reporting of broadband access within the FCC Fixed Broadband Deployment data. FCC data measures where there is broadband available according to internet service providers. Thus, the FCC data provides a picture of claimed availability—what may be possible or what is potentially available at the physical or technological level.

Because the forms are submitted biannually, data is available for June and December of each year beginning December 2014.

The FCC fixed broadband data sets are key to knowing where fixed broadband internet is provided however due to their size (> 8 GB) and complexity many researchers avoid using them. The data cannot be used in its raw form available from the FCC website because there are multiple rows per Census Block per unique broadband provider, technology, and speed. This means there are numerous options for measurement of broadband access which could be chosen to fit a given research question. The raw data has many nuances which can be confusing, so it is vital to understand its format before rolling it up to fit a research question. Additionally, FCC data are at the Census Block level, the smallest unit of Census Geography, which gives greater flexibility. It can be rolled up to any higher Census Geography such as Census Block Group, Census Tract, or County.

This package has functions to direct users to the file download URLs, put the dataset into a SQLite database, and process the data to a smaller level for both the new and old format of data. The processed form of the data is one row per specified census geography, such as census block, and counts the number of distinct internet service providers providing internet at or above the given download and upload speed thresholds, up to 5 speed thresholds combinations can be specified. It is also possible to exclude different broadband technologies or look at specific states.

### Census Geography

The FCC data are reported at the Census Block level using the 15-digit FIPS code. This 15-digits FIPS code allows us to group the data to a bigger geographic level, such as Census Block group or County. The 15-digit FIPS code works as follows:

\* AABBBCCCCCCDEEE  
 - A: state  
 - B: county  
 - C: Census Tract  
 - D: Census Block Group  
 - E: Census Block

Each group is nested within the previous (e.g., Counties are nested within States, meaning FIPS county codes are unique within states, and so on for smaller census geography units). State is the largest and Census Block is the smallest.

### Technology Codes

Within the old FCC Fixed Broadband Deployment Data there are 14 technology codes:

- 10: Asymmetric xDSL  
- 11: ADSL2, ADSL2+  
- 12: VDSL  
- 20: Symmetric xDSL  
- 30: All other copper-wire tech  
- 40: Cable Modem other  
- 41: Cable Modem – DOCSIS 1, 1.1 or 2.0  
- 42: Cable Modem – DOCSIS 3.0  
- 43: Cable Modem – DOCSIS 3.1   
- 50: Fiber to the end user  
- 60: Satellite  
- 70: Terrestrial Fixed Wireless  
- 90: Electric Power Line  
- 0: All Other

Within the new FCC Fixed Broadband Deployment Data there are 9 technology codes:

- 10: Copper Wire  
- 40: Coaxial Cable / HFC  
- 50: Optical Carrier / Fiber to the Premises  
- 60: Geostationary Satellite  
- 61: Non-geostationary Satellite  
- 70: Unlicensed Fixed Wireless  
- 71: Licensed Fixed Wireless  
- 72: Licensed-by-Rule Fixed Wireless  
- 0: Other

Not all technologies are equally effective or reliable. The types of technology are not discussed here, but for more information visit [BroadbandNow](https://broadbandnow.com/research).

The option to process the FCC data without certain technologies came about because in our research projects we excluded satellite and fixed wireless technologies. This is because, according to the Fourteenth Broadband Deployment Report, the FCC broadband deployment data indicates satellite service is available nearly everywhere, however subscription rates are relatively low. The report also denotes that fixed wireless data in the FCC broadband deployment data follows trends similar to satellite, potentially signifying the deployment data overestimates availability of satellite and fixed wireless technologies. Additionally, fixed wireless and satellite technologies are inconsistent, suffering from issues such as weather interference, and delays more than other broadband technologies. These issues imply that both satellite and fixed wireless technology reporting may be especially suspect as a measure of access; in particular when those considerations are made along rural-urban lines.

Another reason for the flexibility allowed with technology codes is that one may wish to focus only on one type of technology, excluding all others, to see where it is available throughout the country or a state.

### Broadband Speeds

In 2024 the speed threshold for adequate broadband service was updated to a 100 Mbps download speed and 20 Mbps upload speed, according to the FCC.

Currently adequate broadband service is considered a download speed of 25 Mbps and upload speed of 3 Mbps. Required speeds as technology needs continue to grow is a topic of ongoing discussion. The consensus is that the average broadband user needs higher download speeds than upload, but this may be an outdated understanding of consumer needs. The FCC Consumer Broadband Speed Guide indicates that a speed threshold of 25/3 Mbps is adequate for activities such as general usage, streaming video, video conferencing, and gaming. However, the FCC Household Broadband Guide indicates that download speeds of more than 25 Mbps may be necessary for households with moderate to high broadband use by 4 or more users or devices at a time. On the other hand, some say that 25/3 Mbps is an outdated definition of broadband and 100/10 Mbps is standard.

With the various speeds in mind, the processing function of this package was written to allow for flexibility of user inputted speed thresholds. 5 different thresholds can be considered in the process function which counts the number of providers providing internet at or above the threshold within the specified census geography region. One must input a vector of download speeds and a vector of upload speeds of equal length. The elements of the vectors are matched for the thresholds.

## Using processFCC

### New FCC data format

To process a data time point in the “new” format you will need functions: avail\_new\_dates() and rollup\_new\_FCC(). The most efficient way to work with the new FCC data is the public data application programming interface (API). Using the API functionality avoids the laborious task of downloading each data file individually from the website.

To use the FCC API and the functions in this package for the “new” data format, it is necessary need to create an FCC User Registration account and then setup an API key. To create an account, go to <https://broadbandmap.fcc.gov/login> and then click “Create an account” in the bottom right of the sign-in box.

Once logged into your account, click on your username in the top right corner. Then select the “Manage API Access” option. From this new page, click the “Generate” button to generate a new API token. Copy the token value and save it in a separate location. You will need this key and your username in avail\_new\_dates() and rollup\_new\_FCC() to access the database. For more information on creating an API key visit the [FCC API Instructions](https://us-fcc.app.box.com/v/bdc-public-data-api-spec).

Once your account and API key is set up. The function avail\_new\_dates() can be used to see available dates in the new FCC data format.

avail\_new\_dates(  
 fcc\_username = "email@location.com",  
 api\_key = "longstringofapicharacters"  
)

If you already know what date you would like to process you can move right to using rollup\_new\_FCC(). The rollup functions count the number of unique broadband providers providing internet at or above the speed threshold combinations, up to 5 speed thresholds combinations can be specified. It is also possible to exclude specific broadband technologies or process only some states.

The arguments of rollup\_new\_FCC() are as follows:

* fcc\_username: username for existing FCC account
* api\_key: user’s unique API key for accessing FCC data. Generated within FCC account.
* get\_year: the year of the FCC data to process.
* get\_month: the month of the FCC data to process: either “Jun” or “Dec”.
* states: A vector of the state(s) abbreviations to include in the final data. The default, NULL, includes all states and territories in the final data set.
* geogr: Character representation of Census geography to summarize the data set at: census block (cb), census block group (cbg), census tract (ct), county (county).
* tech\_exc: Vector of technology codes to exclude from data when rolling up. If you do not wish to exclude any technology codes input NA or c(NA). By default, satellite technologies are excluded.
* thresh\_down: Vector of download speeds thresholds with maximum length of 5. The vector must be the same length as thresh\_up because elements of the vectors will be matched to count the number of internet providers at the given download/upload speed combinations.
* thresh\_up: Vector of download speeds thresholds with maximum length of 5. The vector must be the same length as thresh\_down because elements of the vectors will be matched to count the number of internet providers at the given download/upload speed combinations.
* save\_csv: Logical for whether or not to save the processed data as a CSV.
* wd: filepath representing the working directory where the CSV should be saved. By default, this argument is set to the current working directory which is the file location in a qmd/rmd document or R project.

In the following call to rollup\_new\_FCC(), the June 2022 FCC data is rolled up to the census block level for all states (states = NULL) excluding technology codes 60, and 70. For each census block, the processed data considers 5 threshold combinations for download/upload speeds: 25/3, 25/5, 50/5, 75/10, and 100/100 Mbps. The processed data will count the number of providers within the census block providing broadband at the threshold speeds and excluding the specified technologies. No CSV file is saved with the final dataset.

rollup\_new\_FCC(  
 fcc\_username = "email@location.com",  
 api\_key = "longstringofapicharacters",  
 get\_year = "2022",  
 get\_month = "Jun",  
 states = NULL,  
 geogr = "cb",  
 tech\_exc = c("60", "70"),  
 thresh\_down = c(25, 25, 50, 100, 100),  
 thresh\_up = c(3, 5, 10, 10, 100),  
 save\_csv = FALSE,  
 wd = getwd()  
)

### Old FCC data format

If you wish to use a time point of FCC data from the “old” format you will need to use the following functions: old\_FCC\_links(), csv\_to\_sql\_db(), and rollup\_old\_FCC().

old\_FCC\_links() returns URL(s)for the download website for the unprocessed data set of interest. If year and month are left NULL then the output data set will contain all available years and months for the old format. Set most\_recent to TRUE in order to only get the link for the most recent version of the dataset(s), otherwise all available versions will be output. For example, June 2018 is one time point of the data and there may be multiple versions of one time point as the data is updated.

The following code results in a data set with the URLs for where to download the most recent versions of the June and December 2020 FCC data.

old\_FCC\_links(year = 2020, month = NULL, most\_recent = TRUE)

Based on the FCC data time point you would like to process, copy the link from the old\_FCC\_links() output, go to the website, and download the US - Fixed with Satellite data set under the “Fixed Broadband Deployment Block Data” header. The code should still work if you decide to work with a state-level data set but the code is originally designed to work with the U.S. data set. Once on the website from old\_FCC\_links(), the link for the csv file may take you to dropbox. In this case you will need to click on the 3 dots icon on the top bar of the page and then click download.

Next, the FCC CSV file should be added to a SQLite database. To do this, first, create the database connection with dbConnect(). Here, we name the database fcc.sqlite. You can use dbListTables() to check if any tables already exist in the database if you have used the connection previously. Next, in the call to csv\_to\_sql\_db(), we add the FCC data for June 2020 in Alaska to the SQLite database. The name of the table being created in the SQLite database is fcc\_ak\_2020\_Jun, the default table name would not include the Alaska designation. In csv\_to\_sql\_db(), you must set the arguments csv\_file and con.

fcc\_con <- dbConnect(SQLite(), dbname = "fcc.sqlite")  
dbListTables(fcc\_con)  
  
csv\_to\_sql\_db(csv\_file = "C:/Users/kaile/Downloads/AK-Fixed-Jun2020-v2.csv",   
 con = fcc\_con,   
 new\_tbl\_name = "fcc\_ak\_2020\_Jun",  
 year = 2020, month = "Jun",  
 pre\_process\_size = 1000, chunk\_size = 50000,   
 show\_progress\_bar = TRUE)

Once the data is in the SQLite database, we can then process it using rollup\_old\_FCC(). The output of this function is a data set based on the specifications from the function arguments similar to rollup\_new\_FCC(). The following call to rollup\_old\_FCC() processes the June 2020 Alaska FCC data that was just loaded to the SQLite database. The processed data is rolled up to the Census Tract level excluding technology codes 0, 60, and 70 and looks at 5 threshold combinations for download/upload speeds: 25/3, 25/5, 50/5, 75/10, and 100/100 Mbps. The processed data will count the number of providers within the census tract providing broadband at the thresholds. No CSV file is saved with the final dataset.

In the rollup\_old\_FCC() function, the arguments con and table\_in\_con are the only two differences from rollup\_new\_FCC(): con is a SQLite database connection and table\_in\_con is the table in the con database to process.

processed\_dat <- rollup\_old\_FCC(  
 con = fcc\_con,  
 table\_in\_con = "fcc\_ak\_2020\_Jun",  
 year = 2020, month = "Jun",  
 geogr = "ct",  
 tech\_exc = c("0", "60", "70"),  
 thresh\_down = c(25, 25, 50, 75, 100),  
 thresh\_up = c(3, 5, 5, 10, 100)  
 )  
head(processed\_dat)

If you are done working with the SQLite database, be sure to disconnect from the database with dbDisconnect(). You may also want to remove the sqlite, ZIP, and raw CSV files from your working directory because they are considerably large and take up a lot of space.

dbDisconnect(con)  
file.remove("fcc.sqlite")

## Remarks

Once the FCC fixed broadband data is on a level where there is one row per census geography region it is much easier to work with and opens a world of possibilities, such as calculating the proportion of people within a larger census geography with access to a certain speed or making maps of the number of providers per region. For example, here is a map of the state of Iowa at the Census Block level created using the FCC data. Satellite technology was excluded in rolling up the data to a 25/3 threshold. The provider count at 25/3 Mbps was then converted to a binary variable where 1 indicates access to broadband at 25/3 and 0 indicates no access to internet at 25/3.

