

Baby Boomers Love Their Drugs

Yogindra Raghav

December 10, 2018

I was interested in looking into Baby Boomers usage of drugs after a conversation with one of my friends. This came about during a Genomics class which dealt with Genomic Medicine and Pharmacogenomics. It's a long story how we got to that topic but either way I found a data set to use and even found an article that did some visualizations of said data.

Without further ado, I will be replicating the data graphic about Baby Boomers and drug use from here.

Note: I will do technical discussion of the data wrangling-visualization procedure step-by-step as I replicate the graphic. After creation of the graph, I will explain the context of the data graphic.

ANALYSIS:

Use the `mdsr` and `fivethirtyeight` packages to do this analysis.

```
# load necessary packages

library(mdsr)
library(fivethirtyeight) # package that contains data
library(tibble)
```

To recreate the graph and look at drug use amongst Baby Boomers, we will look at the `drug_use` dataset. This dataset contains usage statistics of 13 drugs in the past 12 months over 17 age groups.

```
# two ways of visualizing data set:
```

```
drug_use
```

```
## # A tibble: 17 x 28
##   age      n alcohol_use alcohol_freq marijuana_use marijuana_freq
##   <ord> <int>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 12      2798         3.9         3         1.1         4
## 2 13      2757         8.5         6         3.4        15
## 3 14      2792        18.1         5         8.7        24
## 4 15      2956        29.2         6        14.5        25
## 5 16      3058        40.1        10        22.5        30
## 6 17      3038        49.3        13        28         36
## 7 18      2469        58.7        24        33.7        52
## 8 19      2223        64.6        36        33.4        60
## 9 20      2271        69.7        48        34         60
## 10 21      2354        83.2        52        33         52
## 11 22-23    4707        84.2        52        28.4        52
## 12 24-25    4591        83.1        52        24.9        60
## 13 26-29    2628        80.7        52        20.8        52
## 14 30-34    2864        77.5        52        16.4        72
## 15 35-49    7391         75         52        10.4        48
## 16 50-64    3923        67.2        52         7.3        52
## 17 65+      2448        49.3        52         1.2        36
## # ... with 22 more variables: cocaine_use <dbl>, cocaine_freq <dbl>,
## #   crack_use <dbl>, crack_freq <dbl>, heroin_use <dbl>,
## #   heroin_freq <dbl>, hallucinogen_use <dbl>, hallucinogen_freq <dbl>,
## #   inhalant_use <dbl>, inhalant_freq <dbl>, pain_releiver_use <dbl>,
```

```
## #   pain_releiver_freq <dbl>, oxycontin_use <dbl>, oxycontin_freq <dbl>,
## #   tranquilizer_use <dbl>, tranquilizer_freq <dbl>, stimulant_use <dbl>,
## #   stimulant_freq <dbl>, meth_use <dbl>, meth_freq <dbl>,
## #   sedative_use <dbl>, sedative_freq <dbl>
```

```
glimpse(drug_use)
```

```
## Observations: 17
## Variables: 28
## $ age          <ord> 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22-...
## $ n            <int> 2798, 2757, 2792, 2956, 3058, 3038, 2469, 2...
## $ alcohol_use  <dbl> 3.9, 8.5, 18.1, 29.2, 40.1, 49.3, 58.7, 64....
## $ alcohol_freq <dbl> 3, 6, 5, 6, 10, 13, 24, 36, 48, 52, 52, 52,...
## $ marijuana_use <dbl> 1.1, 3.4, 8.7, 14.5, 22.5, 28.0, 33.7, 33.4...
## $ marijuana_freq <dbl> 4, 15, 24, 25, 30, 36, 52, 60, 60, 52, 52, ...
## $ cocaine_use  <dbl> 0.1, 0.1, 0.1, 0.5, 1.0, 2.0, 3.2, 4.1, 4.9...
## $ cocaine_freq <dbl> 5.0, 1.0, 5.5, 4.0, 7.0, 5.0, 5.0, 5.5, 8.0...
## $ crack_use    <dbl> 0.0, 0.0, 0.0, 0.1, 0.0, 0.1, 0.4, 0.5, 0.6...
## $ crack_freq   <dbl> NA, 3.0, NA, 9.5, 1.0, 21.0, 10.0, 2.0, 5.0...
## $ heroin_use    <dbl> 0.1, 0.0, 0.1, 0.2, 0.1, 0.1, 0.4, 0.5, 0.9...
## $ heroin_freq   <dbl> 35.5, NA, 2.0, 1.0, 66.5, 64.0, 46.0, 180.0...
## $ hallucinogen_use <dbl> 0.2, 0.6, 1.6, 2.1, 3.4, 4.8, 7.0, 8.6, 7.4...
## $ hallucinogen_freq <dbl> 52, 6, 3, 4, 3, 3, 4, 3, 2, 4, 3, 2, 3, 2, ...
## $ inhalant_use  <dbl> 1.6, 2.5, 2.6, 2.5, 3.0, 2.0, 1.8, 1.4, 1.5...
## $ inhalant_freq <dbl> 19.0, 12.0, 5.0, 5.5, 3.0, 4.0, 4.0, 3.0, 4...
## $ pain_releiver_use <dbl> 2.0, 2.4, 3.9, 5.5, 6.2, 8.5, 9.2, 9.4, 10....
## $ pain_releiver_freq <dbl> 36, 14, 12, 10, 7, 9, 12, 12, 10, 15, 15, 1...
## $ oxycontin_use <dbl> 0.1, 0.1, 0.4, 0.8, 1.1, 1.4, 1.7, 1.5, 1.7...
## $ oxycontin_freq <dbl> 24.5, 41.0, 4.5, 3.0, 4.0, 6.0, 7.0, 7.5, 1...
## $ tranquilizer_use <dbl> 0.2, 0.3, 0.9, 2.0, 2.4, 3.5, 4.9, 4.2, 5.4...
## $ tranquilizer_freq <dbl> 52.0, 25.5, 5.0, 4.5, 11.0, 7.0, 12.0, 4.5,...
## $ stimulant_use  <dbl> 0.2, 0.3, 0.8, 1.5, 1.8, 2.8, 3.0, 3.3, 4.0...
## $ stimulant_freq <dbl> 2.0, 4.0, 12.0, 6.0, 9.5, 9.0, 8.0, 6.0, 12...
## $ meth_use       <dbl> 0.0, 0.1, 0.1, 0.3, 0.3, 0.6, 0.5, 0.4, 0.9...
## $ meth_freq      <dbl> NA, 5.0, 24.0, 10.5, 36.0, 48.0, 12.0, 105....
## $ sedative_use   <dbl> 0.2, 0.1, 0.2, 0.4, 0.2, 0.5, 0.4, 0.3, 0.5...
## $ sedative_freq  <dbl> 13.0, 19.0, 16.5, 30.0, 3.0, 6.5, 10.0, 6.0...
```

Since we need to visualize the percentage of Americans between the ages of 50-64 that took drugs, we can filter the age column in the data set. The visualization of the new dataset is also shown.

```
baby_boomers = drug_use %>% filter(age == "50-64") # filter
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
glimpse(baby_boomers) # view data set
```

```
## Observations: 1
## Variables: 28
## $ age          <ord> 50-64
## $ n            <int> 3923
## $ alcohol_use  <dbl> 67.2
## $ alcohol_freq <dbl> 52
## $ marijuana_use <dbl> 7.3
## $ marijuana_freq <dbl> 52
## $ cocaine_use  <dbl> 0.9
## $ cocaine_freq <dbl> 36
```

```
## $ crack_use      <dbl> 0.4
## $ crack_freq     <dbl> 62
## $ heroin_use      <dbl> 0.1
## $ heroin_freq     <dbl> 41
## $ hallucinogen_use <dbl> 0.3
## $ hallucinogen_freq <dbl> 44
## $ inhalant_use    <dbl> 0.2
## $ inhalant_freq   <dbl> 13.5
## $ pain_releiver_use <dbl> 2.5
## $ pain_releiver_freq <dbl> 12
## $ oxycontin_use   <dbl> 0.4
## $ oxycontin_freq  <dbl> 5
## $ tranquilizer_use <dbl> 1.4
## $ tranquilizer_freq <dbl> 10
## $ stimulant_use    <dbl> 0.3
## $ stimulant_freq   <dbl> 24
## $ meth_use         <dbl> 0.2
## $ meth_freq        <dbl> 30
## $ sedative_use     <dbl> 0.2
## $ sedative_freq    <dbl> 104
```

Now we will go ahead and select for columns in our dataset that are used in the graphic of the article. There are several. Most importantly each drug has two columns. One column is for whether a person has used that drug in the last 12 months. The other column explains the frequency at which they did so. Based on the graphic chosen, we need to visualize the number of people that have used the drug. For this reason, we shall select for columns that deal with drug use for those drugs used in the graphic from the article.

```
baby_boomers = baby_boomers %>% select(marijuana_use, pain_releiver_use, tranquilizer_use,
cocaine_use, crack_use, oxycontin_use, stimulant_use, hallucinogen_use, sedative_use,
inhalant_use, meth_use, heroin_use) # selecting columns

glimpse(baby_boomers) # visualize
```

```
## Observations: 1
## Variables: 12
## $ marijuana_use    <dbl> 7.3
## $ pain_releiver_use <dbl> 2.5
## $ tranquilizer_use <dbl> 1.4
## $ cocaine_use      <dbl> 0.9
## $ crack_use        <dbl> 0.4
## $ oxycontin_use    <dbl> 0.4
## $ stimulant_use    <dbl> 0.3
## $ hallucinogen_use <dbl> 0.3
## $ sedative_use     <dbl> 0.2
## $ inhalant_use     <dbl> 0.2
## $ meth_use         <dbl> 0.2
## $ heroin_use        <dbl> 0.1
```

For the sake of plotting, we need to transpose the table (flip rows and columns). Once this is done we convert to a data frame so that we can add a column for the drug and change column names. We convert this data frame to a tibble and then use the row names (which are the drugs) for a new column. Create a variable with a list of the drugs in the correct order. Once that's done we make a new column with the correct drug names that are used in the final graphic. Rename the percentage frequency column to frequency. Select for the correct drug names and frequency column.

```

baby_boomers = t(baby_boomers) # transpose

baby_boomers = data.frame(baby_boomers) # make into data frame to keep the drug names

baby_boomers = rownames_to_column(as_tibble(baby_boomers))
# convert baby_boomers to tibble and then make drugs into a column

drug_names = c("Marijuana", "Pain reliever", "Tranquilizer", "Cocaine", "Crack",
               "OxyContin", "Stimulant", "Hallucinogen", "Sedative", "Inhalant", "Meth",
               "Heroin") # drug names

baby_boomers$drugs = drug_names # make new column

baby_boomers = rename(baby_boomers, frequency = baby_boomers) # rename column

baby_boomers # show

```

```

## # A tibble: 12 x 3
##   rowname      frequency drugs
##   <chr>          <dbl> <chr>
## 1 marijuana_use      7.3 Marijuana
## 2 pain_releiver_use  2.5 Pain reliever
## 3 tranquilizer_use   1.4 Tranquilizer
## 4 cocaine_use        0.9 Cocaine
## 5 crack_use          0.4 Crack
## 6 oxycontin_use      0.4 OxyContin
## 7 stimulant_use      0.3 Stimulant
## 8 hallucinogen_use   0.3 Hallucinogen
## 9 sedative_use       0.2 Sedative
## 10 inhalant_use      0.2 Inhalant
## 11 meth_use          0.2 Meth
## 12 heroin_use        0.1 Heroin

```

```

baby_boomers = baby_boomers %>% select(drugs, frequency) # select

```

```

baby_boomers

```

```

## # A tibble: 12 x 2
##   drugs      frequency
##   <chr>          <dbl>
## 1 Marijuana      7.3
## 2 Pain reliever  2.5
## 3 Tranquilizer   1.4
## 4 Cocaine        0.9
## 5 Crack          0.4
## 6 OxyContin      0.4
## 7 Stimulant      0.3
## 8 Hallucinogen   0.3
## 9 Sedative       0.2
## 10 Inhalant      0.2
## 11 Meth          0.2
## 12 Heroin        0.1

```

For the sake of the bar graph going in descending order we must use the reorder function first. This will make the `baby_boomers$drugs` column in our `baby_boomers` data set into a factor. Once we have this, we

can start plotting. First we give the data to be plotted to ggplot, then we let it know that we want to map drugs and their frequency. Next we tell it that we want to make a bar graph. In this case, you cannot use the default `geom_bar()` function since it uses `stat_count()` function which does not work with our data. You need to tell it that we want to use the exact numbers in our `baby_boomers$frequency` column to graph the data. For this reason, we use the `geom_bar(stat="identity")` option. For the sake of replicating the original graph, I made sure to tell it to fill the bars with the color red. After this we need to flip the coordinates since we want the drugs to be on the y-axis. The original data graphic in the chosen article is stripped of all titles, backgrounds, axis, etc. The only things that remain are the drug names, a title for the whole plot and the corresponding frequency numbers for each bar. For this reason I gave multiple options to the `theme()` function which hid all of those elements that were unnecessary. I then used the `geom_text()` function to add the frequency numbers correctly to the right of the corresponding bars. Finally I gave it a title using `ggtitle()`. The one important thing though is that I had to use the `\n` character (used commonly in computer science) to make a newline so that the title would be two lines, just like the original graph.

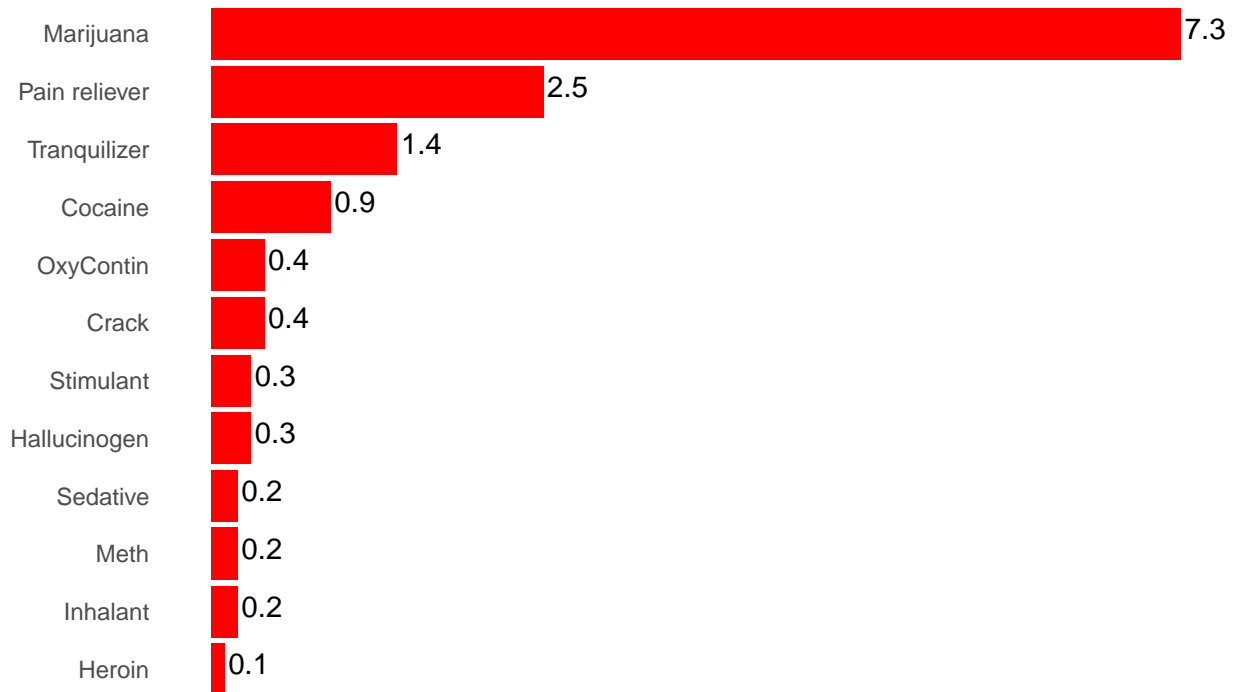
```

baby_boomers$drugs = reorder(baby_boomers$drugs, baby_boomers$frequency) # reorder

ggplot(data = baby_boomers, aes(x = drugs, y = frequency))+
  geom_bar(stat= "identity", fill = "red")+
  coord_flip()+
  theme(axis.line= element_blank(), axis.text.x = element_blank(),
        axis.ticks = element_blank(), axis.title.x = element_blank(),
        axis.title.y =element_blank(),panel.background=element_blank(),
        panel.border=element_blank(),panel.grid.major=element_blank(),
        panel.grid.minor=element_blank(),plot.background=element_blank())+
  geom_text(aes(label = frequency), vjust = .24, hjust = -.1)+
  ggtitle("Percentage of Americans aged 50-64 who said in a 2012 survey that \nthey had
used the following drugs in the past year")

```

Percentage of Americans aged 50–64 who said in a 2012 survey that they had used the following drugs in the past year



CONTEXT:

The original dataset that the article uses for its own visualization purposes comes from the National Survey on Drug Use and Health from the Substance Abuse and Mental Health Data Archive. The authors of the original visualization wanted to see what drugs the “Baby Boomers” are currently taking. This is an interesting question for society since the “Baby Boomer” generation always has the stigma/connotation of being notorious drug users. This was probably why this graphic was chosen. Alas, the whole point was to see what drugs that generation are currently using and what proportion of the generation does so.