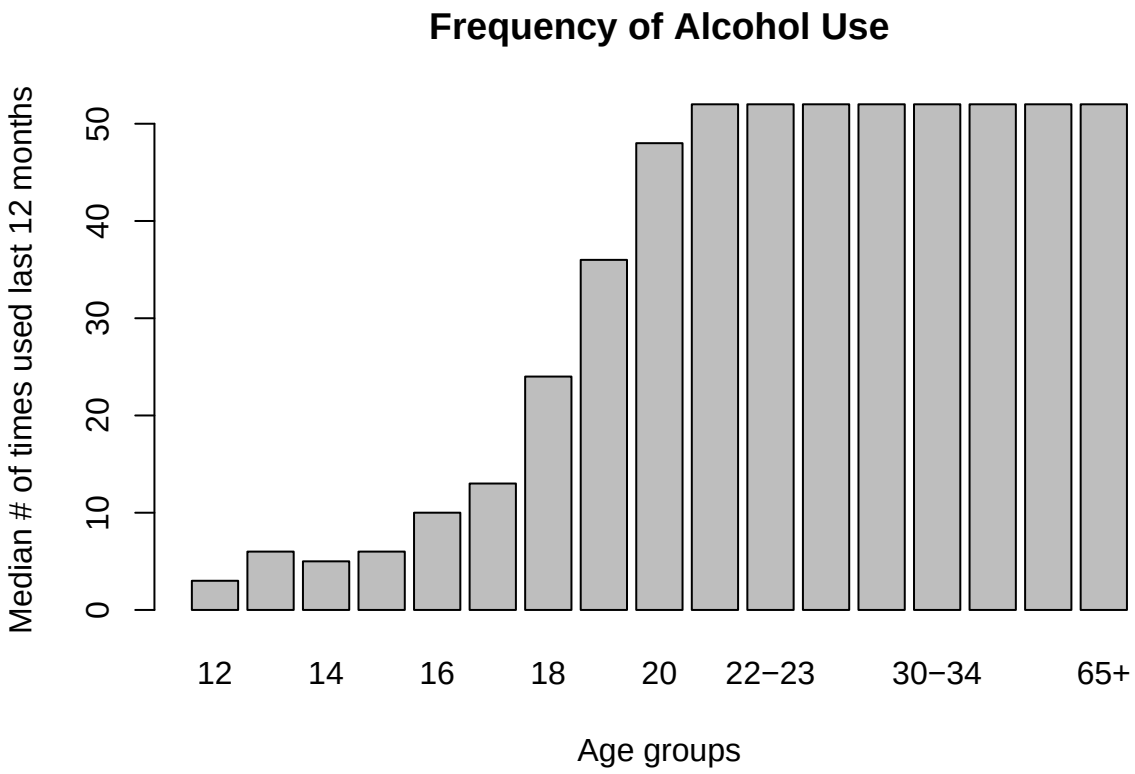
# 1st Pair Assignment

*Chris Kardish and Marie Agosta*

*September 30, 2016*

This is the first pair assignment from Chris Kardish and Marie Agosta. The first data set comes via FiveThirtyEight; the second is core R and is sourced from a second script. First, the FiveThirtyEight stuff. It is a survey of drug use by age group carried out by the US Substance Abuse and Mental Health Services Administration.

There are many drugs to choose from, but for purposes of brevity we'll only cover the descriptive statistics of a few, starting with the lighter stuff.




You notice immediately from the barplot that this particular variable is a bit odd: the median rises steadiliy with age before completely leveling off at 52 at age 21. That's likely a problem with SAMHSA's or FiveThirtyEight's reporting and it will affect the following descriptive statistics.

Mean:

```
## [1] 33.35294
```

Median:

```
## [1] 48
```

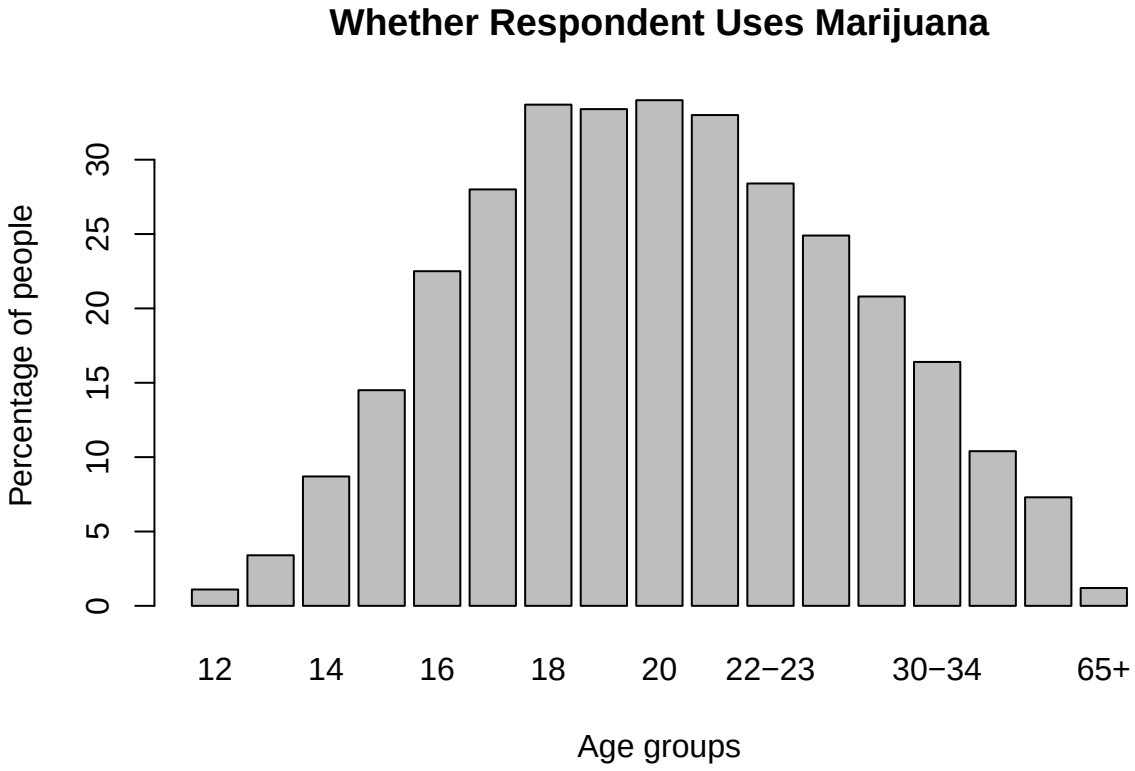
Standard Deviation:

```
## [1] 21.31883
```

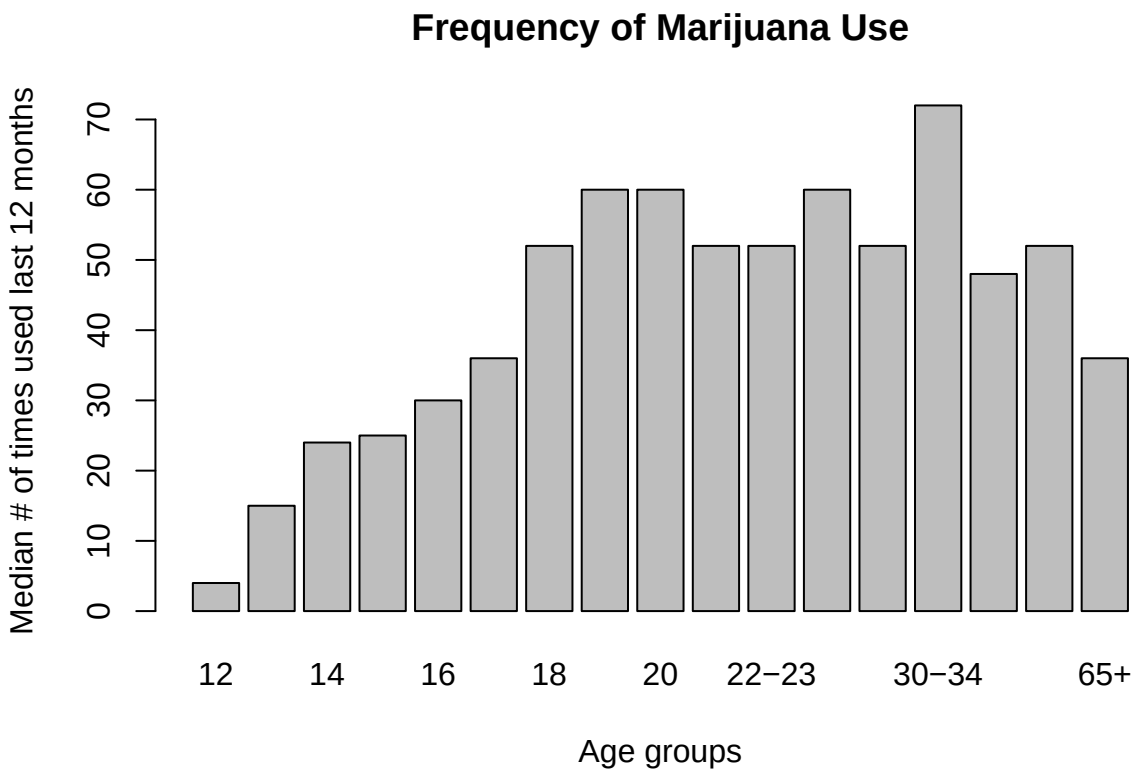
Range:

```
## [1]  3 52
```

Moving on to marijuana. Here's a plot for percentage of people in a given age group who have used marijuana in the past 12 months.




Now here's a barplot for marijuana use by age, defined as frequency of use in the last 12 months.This of course only applies to people who identify as users.




Initial visual hypothesis: as you age you're less likely to be a marijuana user, but if you do remain a user into your 30s you're likely to smoke as much as the kids or even more. But obviously this idea would have to be tested for statistical significance.

Let's close things out by answering the following question: is the average number of heroin users larger than the average number of meth users? Let's find out!

Percentage of those in any age group who used heroin in the past 12 months:

```
## [1] 0.3529412
```

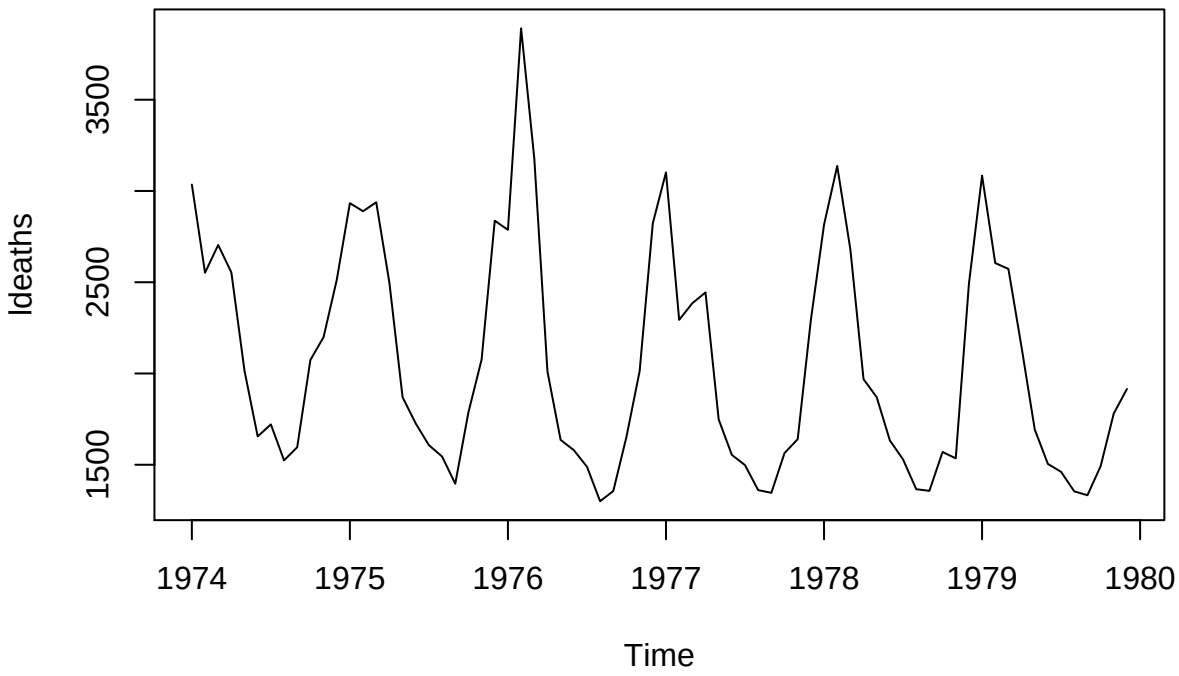
Percentage of those in any age group who used meth in the past 12 months:

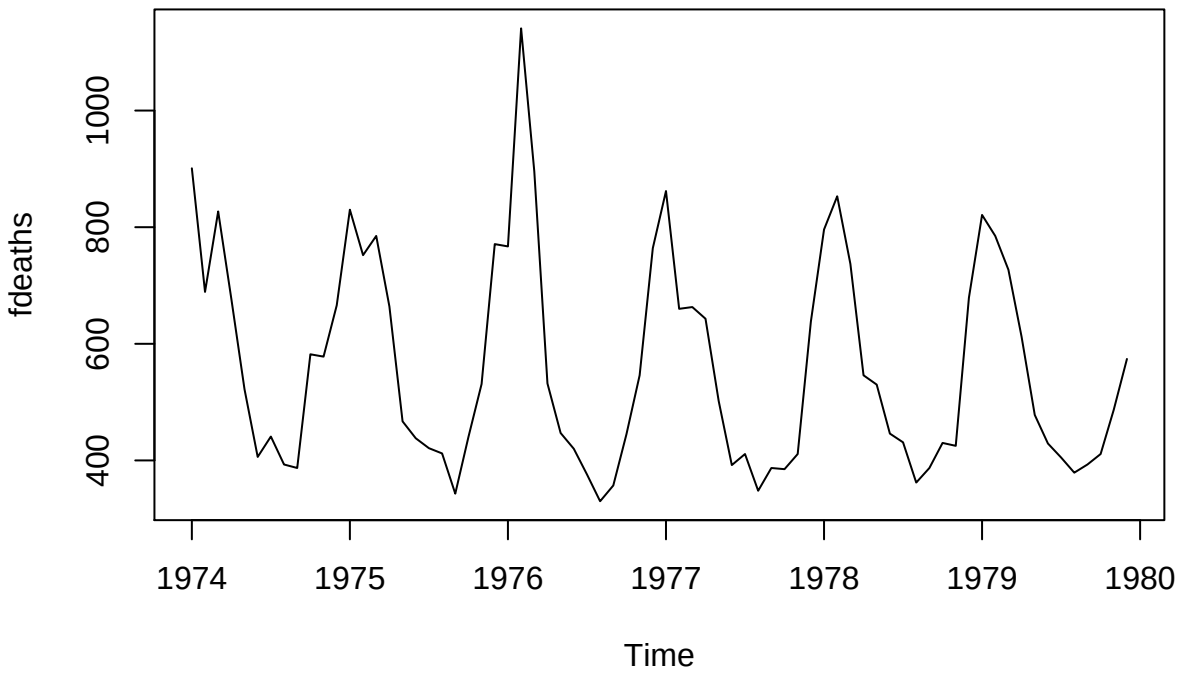
```
## [1] 0.3823529
```

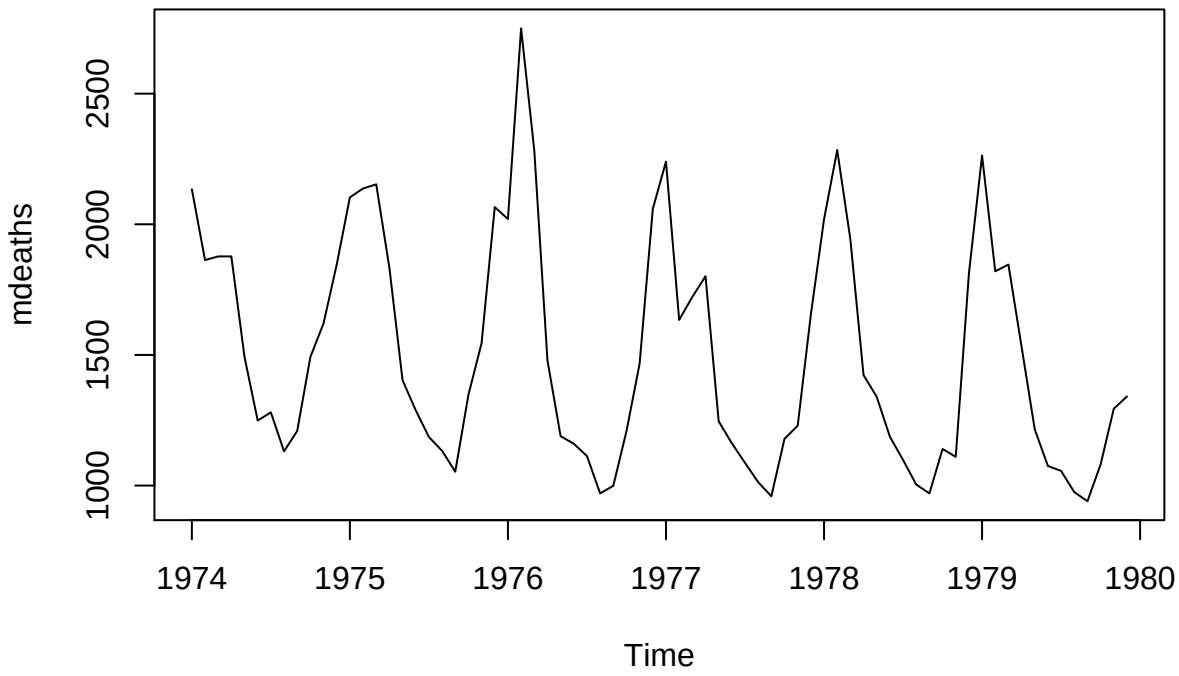
It appears meth remains slightly more popular, at least from this vantage point.

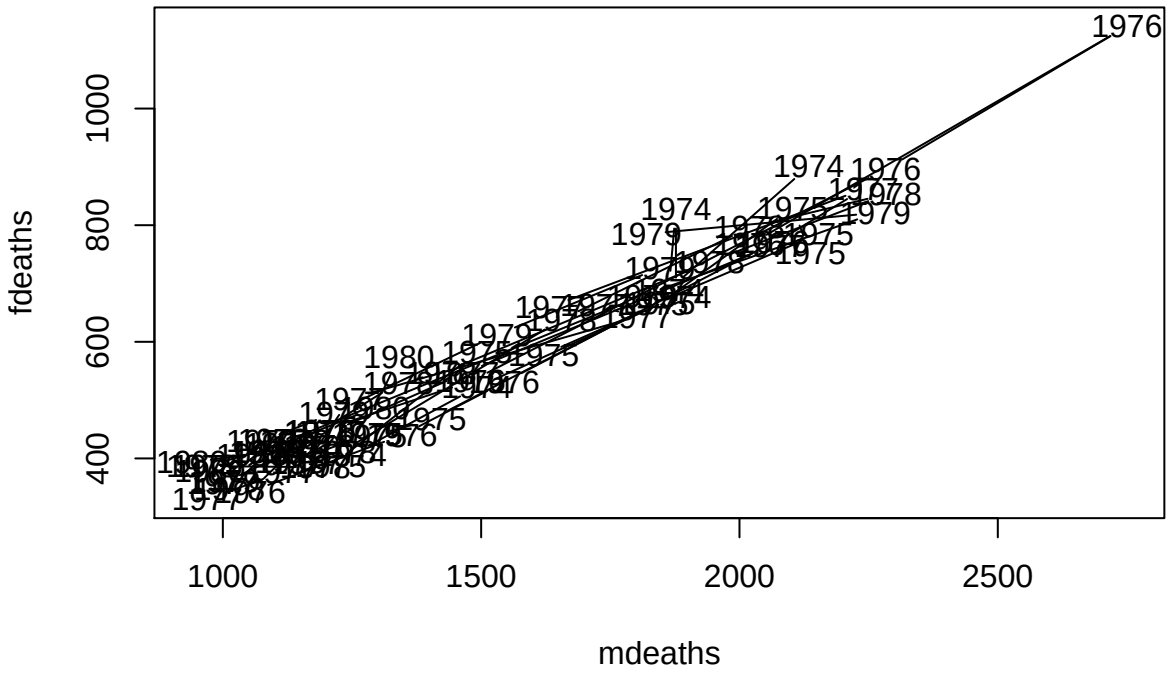
Lastly, we call up another R script that also explores descriptive statistics using the "source" function.Doing so also brings the underlying data set for it (deaths from various lung diseases in the UK)

```
source("MCAgosta.R")
```


ldeaths
1500
2500
3500
1974
1975
1976
1977
1978
1979
1980
Time


fdeaths
400
600
800
1000
1974
1975
1976
1977
1978
1979
1980
Time


mdeaths
1000
1500
2000
2500
1974
1975
1976
1977
1978
1979
1980
Time


fdeaths
mdeaths
400
600
800
1000
1000
1500
2000
2500
1976
1974
1979
1975