Capstone 2 Data Only

Suicide and Smartphones Data Compilation

This document is created simply to show how the dataset named "wide" was created that was used in the "Suicide and Smartphones" report. It does not have the sections and layout that are required in the second Harvard "Data Science" series capstone project 2.

Installing R packages needed

```
# installing the R packages needed
if(!require(readxl)) install.packages("readxl", repos = "http://cran.us.r-project.org")
```

```
## Warning: package 'readxl' was built under R version 3.5.3
```

```
if(!require(tidyverse)) install.packages("tidyverse", repos = "http://cran.us.r-project.org")
```

```
## Warning: package 'tidyverse' was built under R version 3.5.3
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
## Warning: package 'readr' was built under R version 3.5.3
```

```
## Warning: package 'purrr' was built under R version 3.5.3
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
## Warning: package 'forcats' was built under R version 3.5.3
```

```
if(!require(lubridate)) install.packages("lubridate", repos = "http://cran.us.r-project.org")
```

```
## Warning: package 'lubridate' was built under R version 3.5.3
```

Importing the data

```
# importing data
getwd()
```

```
## [1] "C:/Users/aausp/Documents/R/Capstone 2 - GitHub/Upload"
```

```
Study_1970_2002 <- read_excel("Study 1970-2002 suicide rate by age group.xlsx")
n15_19 <- read_excel("CDC suicide rate ages 15-19 1999 to 2017.xlsx")
n10_14 <- read_excel("CDC Suicide rate ages 10-14 1999 to 2017.xlsx")
n10_24 <- read_excel("CDC suicide rate ages 10-24 1999 to 2017.xlsx")
n20_24 <- read_excel("CDC Suicide rate ages 20-24 1999 to 2017.xlsx")
Census_compint <- read_excel("computer and internet use 1984-2009.xlsx")
Pew_SM <- read_excel("Pew Social Media Use 2.xlsx")
all_suicides <- read_excel("CDC all age suicide rate 1999 to 2017.xlsx")
cell_own <- read_excel("Cell smart phone ownership.xlsx")

# importing stock market data

Dow <- read_excel("Stock Market Index Dow Data.xlsx")
```

```
SP <- read_excel("Stock Market Index SP 500 Data.xlsx")
nasdaq <- read_excel("Stock Market Index nasdaq data.xlsx")
```

Stock Market Data Compilation

```
# examining the imported stock market data

head(Dow)
```

```
## # A tibble: 6 x 7
##   Date                 Open  High   Low Close `Adj Close`    Volume
##   <dttm>              <dbl> <dbl> <dbl> <dbl>       <dbl>     <dbl>
## 1 1985-01-01 00:00:00 1278. 1305. 1267. 1287.       1287.  44450000
## 2 1985-02-01 00:00:00 1277. 1308. 1264. 1284.       1284. 207300000
## 3 1985-03-01 00:00:00 1285. 1310. 1243. 1267.       1267. 201050000
## 4 1985-04-01 00:00:00 1265. 1290. 1246. 1258.       1258. 187110000
## 5 1985-05-01 00:00:00 1257. 1321. 1236. 1315.       1315. 242250000
## 6 1985-06-01 00:00:00 1321. 1341. 1285. 1335.       1335. 205340000
```

```
head(SP)
```

```
## # A tibble: 6 x 7
##   Date                 Open  High   Low Close `Adj Close`    Volume
##   <dttm>              <dbl> <dbl> <dbl> <dbl>       <dbl>     <dbl>
## 1 1971-03-01 00:00:00  96.8  102.  96.1 100.        100.  389880000
## 2 1971-04-01 00:00:00 100.   106.  99.6 104.        104.  401580000
## 3 1971-05-01 00:00:00 104.   104.  98.7  99.6        99.6 307360000
## 4 1971-06-01 00:00:00  99.6  102.  96.9  98.7        98.7 303590000
## 5 1971-07-01 00:00:00  99.2  102.  95.1  95.6        95.6 265240000
## 6 1971-08-01 00:00:00  95.6  102.  92.8  99.0        99.0 320520000
```

```
head(nasdaq)
```

```
## # A tibble: 6 x 7
##   Date                 Open  High   Low Close `Adj Close` Volume
##   <dttm>              <dbl> <dbl> <dbl> <dbl>       <dbl>  <dbl>
## 1 1971-02-01 00:00:00 100    102.  99.7 101.        101.      0
## 2 1971-03-01 00:00:00 102.   106. 102.  106.        106.      0
## 3 1971-04-01 00:00:00 106.   112. 106.  112.        112.      0
## 4 1971-05-01 00:00:00 111.   112. 108.  108.        108.      0
## 5 1971-06-01 00:00:00 109.   111. 105.  108.        108.      0
## 6 1971-07-01 00:00:00 108.   111. 105.  105.        105.      0
```

```
# keeping only the necessary columns for the stock market data

Dow <- Dow[,c(1,6)]
nasdaq <- nasdaq[,c(1,6)]
SP <- SP[,c(1,6)]

# examining the new data frames

head(Dow)
```

```
## # A tibble: 6 x 2
##   Date                `Adj Close`
##   <dttm>                    <dbl>
```

```
## 1 1985-01-01 00:00:00        1287.
## 2 1985-02-01 00:00:00        1284.
## 3 1985-03-01 00:00:00        1267.
## 4 1985-04-01 00:00:00        1258.
## 5 1985-05-01 00:00:00        1315.
## 6 1985-06-01 00:00:00        1335.
```

```r
head(SP)
```

```
## # A tibble: 6 x 2
##   Date                `Adj Close`
##   <dttm>                    <dbl>
## 1 1971-03-01 00:00:00        100.
## 2 1971-04-01 00:00:00        104.
## 3 1971-05-01 00:00:00         99.6
## 4 1971-06-01 00:00:00         98.7
## 5 1971-07-01 00:00:00         95.6
## 6 1971-08-01 00:00:00         99.0
```

```r
head(nasdaq)
```

```
## # A tibble: 6 x 2
##   Date                `Adj Close`
##   <dttm>                    <dbl>
## 1 1971-02-01 00:00:00        101.
## 2 1971-03-01 00:00:00        106.
## 3 1971-04-01 00:00:00        112.
## 4 1971-05-01 00:00:00        108.
## 5 1971-06-01 00:00:00        108.
## 6 1971-07-01 00:00:00        105.
```

```r
# creating a column for the year of the Date column

Dow <- Dow %>% mutate(Year = year(Date))
SP <- SP %>% mutate(Year = year(Date))
nasdaq <- nasdaq %>% mutate(Year = year(Date))

# examining the new market data frames

Dow
```

```
## # A tibble: 420 x 3
##    Date                `Adj Close`  Year
##    <dttm>                    <dbl> <dbl>
##  1 1985-01-01 00:00:00        1287.  1985
##  2 1985-02-01 00:00:00        1284.  1985
##  3 1985-03-01 00:00:00        1267.  1985
##  4 1985-04-01 00:00:00        1258.  1985
##  5 1985-05-01 00:00:00        1315.  1985
##  6 1985-06-01 00:00:00        1335.  1985
##  7 1985-07-01 00:00:00        1347.  1985
##  8 1985-08-01 00:00:00        1334.  1985
##  9 1985-09-01 00:00:00        1329.  1985
## 10 1985-10-01 00:00:00        1374.  1985
## # ... with 410 more rows
```

```
SP
```

```
## # A tibble: 586 x 3
##    Date                `Adj Close`  Year
##    <dttm>                    <dbl> <dbl>
##  1 1971-03-01 00:00:00       100.   1971
##  2 1971-04-01 00:00:00       104.   1971
##  3 1971-05-01 00:00:00        99.6  1971
##  4 1971-06-01 00:00:00        98.7  1971
##  5 1971-07-01 00:00:00        95.6  1971
##  6 1971-08-01 00:00:00        99.0  1971
##  7 1971-09-01 00:00:00        98.3  1971
##  8 1971-10-01 00:00:00        94.2  1971
##  9 1971-11-01 00:00:00        94.0  1971
## 10 1971-12-01 00:00:00       102.   1971
## # ... with 576 more rows
```

```
nasdaq
```

```
## # A tibble: 587 x 3
##    Date                `Adj Close`  Year
##    <dttm>                    <dbl> <dbl>
##  1 1971-02-01 00:00:00       101.   1971
##  2 1971-03-01 00:00:00       106.   1971
##  3 1971-04-01 00:00:00       112.   1971
##  4 1971-05-01 00:00:00       108.   1971
##  5 1971-06-01 00:00:00       108.   1971
##  6 1971-07-01 00:00:00       105.   1971
##  7 1971-08-01 00:00:00       108.   1971
##  8 1971-09-01 00:00:00       109.   1971
##  9 1971-10-01 00:00:00       105.   1971
## 10 1971-11-01 00:00:00       104.   1971
## # ... with 577 more rows
```

```
# keeping only the year and adj close columns

Dow <- Dow[,c(2,3)]
SP <- SP[,c(2,3)]
nasdaq <- nasdaq[,c(2,3)]

Dow
```

```
## # A tibble: 420 x 2
##    `Adj Close`  Year
##          <dbl> <dbl>
##  1       1287.  1985
##  2       1284.  1985
##  3       1267.  1985
##  4       1258.  1985
##  5       1315.  1985
##  6       1335.  1985
##  7       1347.  1985
##  8       1334.  1985
##  9       1329.  1985
## 10       1374.  1985
## # ... with 410 more rows
```

```
SP
```

```
## # A tibble: 586 x 2
##    `Adj Close`  Year
##          <dbl> <dbl>
##  1       100.   1971
##  2       104.   1971
##  3        99.6  1971
##  4        98.7  1971
##  5        95.6  1971
##  6        99.0  1971
##  7        98.3  1971
##  8        94.2  1971
##  9        94.0  1971
## 10       102.   1971
## # ... with 576 more rows
```

```
nasdaq
```

```
## # A tibble: 587 x 2
##    `Adj Close`  Year
##          <dbl> <dbl>
##  1       101.   1971
##  2       106.   1971
##  3       112.   1971
##  4       108.   1971
##  5       108.   1971
##  6       105.   1971
##  7       108.   1971
##  8       109.   1971
##  9       105.   1971
## 10       104.   1971
## # ... with 577 more rows
```

```r
# making these data frames all numeric

SP$'Adj Close' <- as.numeric(SP$'Adj Close')
Dow$'Adj Close' <- as.numeric(Dow$'Adj Close')
nasdaq$'Adj Close' <- as.numeric(nasdaq$'Adj Close')

SP
```

```
## # A tibble: 586 x 2
##    `Adj Close`  Year
##          <dbl> <dbl>
##  1       100.   1971
##  2       104.   1971
##  3        99.6  1971
##  4        98.7  1971
##  5        95.6  1971
##  6        99.0  1971
##  7        98.3  1971
##  8        94.2  1971
##  9        94.0  1971
## 10       102.   1971
## # ... with 576 more rows
```

```
Dow
```

```
## # A tibble: 420 x 2
##     `Adj Close`  Year
##           <dbl> <dbl>
##  1       1287.   1985
##  2       1284.   1985
##  3       1267.   1985
##  4       1258.   1985
##  5       1315.   1985
##  6       1335.   1985
##  7       1347.   1985
##  8       1334.   1985
##  9       1329.   1985
## 10       1374.   1985
## # ... with 410 more rows
```

```
nasdaq
```

```
## # A tibble: 587 x 2
##     `Adj Close`  Year
##           <dbl> <dbl>
##  1        101.   1971
##  2        106.   1971
##  3        112.   1971
##  4        108.   1971
##  5        108.   1971
##  6        105.   1971
##  7        108.   1971
##  8        109.   1971
##  9        105.   1971
## 10        104.   1971
## # ... with 577 more rows
```

```r
# grouping by year and summarizing the close amounts

Dow1 <- Dow %>%
  group_by(Year) %>%
  summarize(Avg_close = mean(`Adj Close`))
Dow1
```

```
## # A tibble: 35 x 2
##     Year Avg_close
##    <dbl>     <dbl>
##  1  1985     1346.
##  2  1986     1815.
##  3  1987     2273.
##  4  1988     2077.
##  5  1989     2536.
##  6  1990     2662.
##  7  1991     2964.
##  8  1992     3296.
##  9  1993     3538.
## 10  1994     3793.
## # ... with 25 more rows
```

```r
SP1 <- SP %>%
  group_by(Year) %>%
  summarize(Avg_close = mean(`Adj Close`))
SP1
```

```
## # A tibble: 49 x 2
##     Year Avg_close
##    <dbl>     <dbl>
##  1  1971      98.6
##  2  1972     110.
##  3  1973     107.
##  4  1974      81.5
##  5  1975      87.1
##  6  1976     103.
##  7  1977      97.5
##  8  1978      95.5
##  9  1979     103.
## 10  1980     120.
## # ... with 39 more rows
```

```r
nasdaq1 <- nasdaq %>%
  group_by(Year) %>%
  summarize(Avg_close = mean(`Adj Close`))
nasdaq1
```

```
## # A tibble: 49 x 2
##     Year Avg_close
##    <dbl>     <dbl>
##  1  1971     107.
##  2  1972     129.
##  3  1973     108.
##  4  1974      75.0
##  5  1975      78.1
##  6  1976      90.7
##  7  1977      98.5
##  8  1978     117.
##  9  1979     138.
## 10  1980     170.
## # ... with 39 more rows
```

```r
# combining markets datasets

colnames(Dow1) <- c("Year", "Dow_close")
colnames(SP1) <- c("Year", "SP_Close")
colnames(nasdaq1) <- c("Year", "nasdaq_Close")

Dow1
```

```
## # A tibble: 35 x 2
##     Year Dow_close
##    <dbl>     <dbl>
##  1  1985     1346.
##  2  1986     1815.
##  3  1987     2273.
##  4  1988     2077.
##  5  1989     2536.
```

```
## 6  1990       2662.
## 7  1991       2964.
## 8  1992       3296.
## 9  1993       3538.
## 10 1994       3793.
## # ... with 25 more rows
```

SP1

```
## # A tibble: 49 x 2
##    Year SP_Close
##   <dbl>   <dbl>
## 1  1971    98.6
## 2  1972   110.
## 3  1973   107.
## 4  1974    81.5
## 5  1975    87.1
## 6  1976   103.
## 7  1977    97.5
## 8  1978    95.5
## 9  1979   103.
## 10 1980   120.
## # ... with 39 more rows
```

nasdaq1

```
## # A tibble: 49 x 2
##    Year nasdaq_Close
##   <dbl>        <dbl>
## 1  1971         107.
## 2  1972         129.
## 3  1973         108.
## 4  1974          75.0
## 5  1975          78.1
## 6  1976          90.7
## 7  1977          98.5
## 8  1978         117.
## 9  1979         138.
## 10 1980         170.
## # ... with 39 more rows
```

```r
# joining the data frames to one named "markets"

markets <- inner_join(Dow1, SP1, by = "Year")
markets <- inner_join(markets, nasdaq1, by = "Year")
markets
```

```
## # A tibble: 35 x 4
##    Year Dow_close SP_Close nasdaq_Close
##   <dbl>     <dbl>    <dbl>        <dbl>
## 1  1985     1346.     189.         293.
## 2  1986     1815.     239.         369.
## 3  1987     2273.     286.         400.
## 4  1988     2077.     268.         376.
## 5  1989     2536.     326.         440.
## 6  1990     2662.     333.         404.
## 7  1991     2964.     382.         502.
```

```
##  8  1992     3296.     417.        604.
##  9  1993     3538.     453.        720.
## 10  1994     3793.     461.        754.
## # ... with 25 more rows
```

```r
# creating new column with mean of three market columns

markets1 <- markets %>% mutate(Mkt_close = (Dow_close + SP_Close + nasdaq_Close)/3)
markets1
```

```
## # A tibble: 35 x 5
##     Year Dow_close SP_Close nasdaq_Close Mkt_close
##    <dbl>     <dbl>    <dbl>        <dbl>     <dbl>
##  1  1985     1346.     189.         293.      609.
##  2  1986     1815.     239.         369.      808.
##  3  1987     2273.     286.         400.      986.
##  4  1988     2077.     268.         376.      907.
##  5  1989     2536.     326.         440.     1101.
##  6  1990     2662.     333.         404.     1133.
##  7  1991     2964.     382.         502.     1283.
##  8  1992     3296.     417.         604.     1439.
##  9  1993     3538.     453.         720.     1570.
## 10  1994     3793.     461.         754.     1669.
## # ... with 25 more rows
```

```r
# selecting just the year and average close in a new dataframe

markets <- markets1[,c(1,5)]
markets
```

```
## # A tibble: 35 x 2
##     Year Mkt_close
##    <dbl>     <dbl>
##  1  1985      609.
##  2  1986      808.
##  3  1987      986.
##  4  1988      907.
##  5  1989     1101.
##  6  1990     1133.
##  7  1991     1283.
##  8  1992     1439.
##  9  1993     1570.
## 10  1994     1669.
## # ... with 25 more rows
```

```r
# rounding the summary column

markets$Mkt_close <- round(markets$Mkt_close, 0)
markets
```

```
## # A tibble: 35 x 2
##     Year Mkt_close
##    <dbl>     <dbl>
##  1  1985       609
##  2  1986       808
##  3  1987       986
##  4  1988       907
```

```
## 5   1989       1101
## 6   1990       1133
## 7   1991       1283
## 8   1992       1439
## 9   1993       1570
## 10  1994       1669
## # ... with 25 more rows
```

```r
# renaming the second column to "Markets"

colnames(markets) <- c("Year", "Markets")
markets
```

```
## # A tibble: 35 x 2
##       Year Markets
##      <dbl>   <dbl>
## 1   1985     609
## 2   1986     808
## 3   1987     986
## 4   1988     907
## 5   1989    1101
## 6   1990    1133
## 7   1991    1283
## 8   1992    1439
## 9   1993    1570
## 10  1994    1669
## # ... with 25 more rows
```

Combining other data

```r
# combine 1999-2017 15-19 and 20-24 suicide rate data

wide <- full_join(n15_19, n20_24, by = "Year")
colnames(wide) <- c("Year", "15-19_rate", "20-24_rate")
wide
```

```
## # A tibble: 18 x 3
##       Year `15-19_rate` `20-24_rate`
##      <dbl>        <dbl>        <dbl>
## 1   2000          8          12.5
## 2   2001          7.9        11.9
## 3   2002          7.3        12.3
## 4   2003          7.2        12.1
## 5   2004          8.1        12.5
## 6   2005          7.5        12.4
## 7   2006          7.1        12.5
## 8   2007          6.7        12.6
## 9   2008          7.2        12.7
## 10  2009          7.5        12.6
## 11  2010          7.5        13.6
## 12  2011          8.3        13.6
## 13  2012          8.3        13.7
## 14  2013          8.3        13.7
## 15  2014          8.7        14.2
## 16  2015          9.8        15.1
```

```
## 17  2016          10            16.1
## 18  2017          11.8          17
```

```r
# add in 10-14 rate

wide <- full_join(wide, n10_14, by = "Year")
colnames(wide) <- c("Year", "15-19_rate", "20-24_rate", "10-14_rate")

# add in cell and smart phone ownership

wide <- full_join(wide, cell_own)
wide
```

```
## # A tibble: 50 x 6
##     Year `15-19_rate` `20-24_rate` `10-14_rate` Owns_Cellphone
##     <dbl>        <dbl>        <dbl>        <dbl> <chr>
##  1  2000            8         12.5          1.5 NA
##  2  2001          7.9         11.9          1.3 NA
##  3  2002          7.3         12.3          1.2 0.62
##  4  2003          7.2         12.1          1.1 NA
##  5  2004          8.1         12.5          1.3 0.65
##  6  2005          7.5         12.4          1.3 0.66
##  7  2006          7.1         12.5          1   NA
##  8  2007          6.7         12.6          0.9 0.75
##  9  2008          7.2         12.7          1   0.8
## 10  2009          7.5         12.6          1.3 0.84
## # ... with 40 more rows, and 1 more variable: Owns_Smartphone <chr>
```

```r
# add in 1970-2002 study

wide <- full_join(wide, Study_1970_2002)
wide
```

```
## # A tibble: 50 x 10
##     Year `15-19_rate` `20-24_rate` `10-14_rate` Owns_Cellphone
##     <dbl>        <dbl>        <dbl>        <dbl> <chr>
##  1  2000            8         12.5          1.5 NA
##  2  2001          7.9         11.9          1.3 NA
##  3  2002          7.3         12.3          1.2 0.62
##  4  2003          7.2         12.1          1.1 NA
##  5  2004          8.1         12.5          1.3 0.65
##  6  2005          7.5         12.4          1.3 0.66
##  7  2006          7.1         12.5          1   NA
##  8  2007          6.7         12.6          0.9 0.75
##  9  2008          7.2         12.7          1   0.8
## 10  2009          7.5         12.6          1.3 0.84
## # ... with 40 more rows, and 5 more variables: Owns_Smartphone <chr>,
## #   `15-24` <dbl>, `25-44` <dbl>, `45-64` <dbl>, `<U+F0B3>65` <dbl>
```

```r
# arrange wide by year

wide <- arrange(wide, Year)
wide
```

```
## # A tibble: 50 x 10
##     Year `15-19_rate` `20-24_rate` `10-14_rate` Owns_Cellphone
##     <dbl>        <dbl>        <dbl>        <dbl> <chr>
```

```
## 1   1970           NA          NA         NA NA
## 2   1971           NA          NA         NA NA
## 3   1972           NA          NA         NA NA
## 4   1973           NA          NA         NA NA
## 5   1974           NA          NA         NA NA
## 6   1975           NA          NA         NA NA
## 7   1976           NA          NA         NA NA
## 8   1977           NA          NA         NA NA
## 9   1978           NA          NA         NA NA
## 10  1979           NA          NA         NA NA
## # ... with 40 more rows, and 5 more variables: Owns_Smartphone <chr>,
## #   `15-24` <dbl>, `25-44` <dbl>, `45-64` <dbl>, `<U+F0B3>65` <dbl>
```

```r
# adjusting Pew social media to be usable

Pew_SM
```

```
## # A tibble: 30 x 6
##    `Survey Date`        ...2  ...3  ...4  ...5  `Social Media Use`
##    <dttm>              <lgl> <lgl> <lgl> <lgl>              <dbl>
## 1  2018-01-10 00:00:00 NA    NA    NA    NA                  0.69
## 2  2010-01-19 00:00:00 NA    NA    NA    NA                  0.43
## 3  2014-01-26 00:00:00 NA    NA    NA    NA                  0.62
## 4  2010-11-24 00:00:00 NA    NA    NA    NA                  0.45
## 5  2010-11-28 00:00:00 NA    NA    NA    NA                  0.46
## 6  2016-11-06 00:00:00 NA    NA    NA    NA                  0.69
## 7  2008-12-20 00:00:00 NA    NA    NA    NA                  0.26
## 8  2010-12-21 00:00:00 NA    NA    NA    NA                  0.47
## 9  2009-12-27 00:00:00 NA    NA    NA    NA                  0.42
## 10 2008-12-04 00:00:00 NA    NA    NA    NA                  0.27
## # ... with 20 more rows
```

```r
Pew_SM <- Pew_SM[,c(1,6)]
```

```r
colnames(Pew_SM) <- c("Survey_Date", "Avg_SM_Use")
Pew_SM$Survey_Date <- as_date(Pew_SM$Survey_Date)
class(Pew_SM$Survey_Date)
```

```
## [1] "Date"
```

```r
Pew_SM
```

```
## # A tibble: 30 x 2
##    Survey_Date Avg_SM_Use
##    <date>           <dbl>
## 1  2018-01-10        0.69
## 2  2010-01-19        0.43
## 3  2014-01-26        0.62
## 4  2010-11-24        0.45
## 5  2010-11-28        0.46
## 6  2016-11-06        0.69
## 7  2008-12-20        0.26
## 8  2010-12-21        0.47
## 9  2009-12-27        0.42
## 10 2008-12-04        0.27
## # ... with 20 more rows
```

```
Pew_SM <- Pew_SM %>% mutate(Year = year(Survey_Date))
Pew_SM
```

```
## # A tibble: 30 x 3
##    Survey_Date Avg_SM_Use   Year
##    <date>           <dbl> <dbl>
##  1 2018-01-10        0.69  2018
##  2 2010-01-19        0.43  2010
##  3 2014-01-26        0.62  2014
##  4 2010-11-24        0.45  2010
##  5 2010-11-28        0.46  2010
##  6 2016-11-06        0.69  2016
##  7 2008-12-20        0.26  2008
##  8 2010-12-21        0.47  2010
##  9 2009-12-27        0.42  2009
## 10 2008-12-04        0.27  2008
## # ... with 20 more rows
```

```
Pew_SM1 <- Pew_SM[,c(2,3)]
Pew_SM1
```

```
## # A tibble: 30 x 2
##    Avg_SM_Use   Year
##         <dbl> <dbl>
##  1       0.69  2018
##  2       0.43  2010
##  3       0.62  2014
##  4       0.45  2010
##  5       0.46  2010
##  6       0.69  2016
##  7       0.26  2008
##  8       0.47  2010
##  9       0.42  2009
## 10       0.27  2008
## # ... with 20 more rows
```

```
# group by year and average rate

Pew_SM2 <- Pew_SM1 %>% group_by(Year) %>%
  summarize(Avg_SM_Use = mean(Avg_SM_Use))

Pew_SM2
```

```
## # A tibble: 13 x 2
##     Year Avg_SM_Use
##    <dbl>      <dbl>
##  1  2005      0.065
##  2  2006      0.11
##  3  2008      0.25
##  4  2009      0.383
##  5  2010      0.458
##  6  2011      0.5
##  7  2012      0.553
##  8  2013      0.613
##  9  2014      0.62
```

```
## 10   2015       0.65
## 11   2016       0.69
## 12   2018       0.69
## 13   2019       0.72
```

```r
# add social media data to the wide dataset

wide <- full_join(wide, Pew_SM2)
wide
```

```
## # A tibble: 50 x 11
##     Year `15-19_rate` `20-24_rate` `10-14_rate` Owns_Cellphone
##    <dbl>        <dbl>        <dbl>        <dbl> <chr>
##  1  1970           NA           NA           NA NA
##  2  1971           NA           NA           NA NA
##  3  1972           NA           NA           NA NA
##  4  1973           NA           NA           NA NA
##  5  1974           NA           NA           NA NA
##  6  1975           NA           NA           NA NA
##  7  1976           NA           NA           NA NA
##  8  1977           NA           NA           NA NA
##  9  1978           NA           NA           NA NA
## 10  1979           NA           NA           NA NA
## # ... with 40 more rows, and 6 more variables: Owns_Smartphone <chr>,
## #   `15-24` <dbl>, `25-44` <dbl>, `45-64` <dbl>, `<U+F0B3>65` <dbl>,
## #   Avg_SM_Use <dbl>
```

```r
# add in market data

wide <- full_join(wide, markets)
wide
```

```
## # A tibble: 50 x 12
##     Year `15-19_rate` `20-24_rate` `10-14_rate` Owns_Cellphone
##    <dbl>        <dbl>        <dbl>        <dbl> <chr>
##  1  1970           NA           NA           NA NA
##  2  1971           NA           NA           NA NA
##  3  1972           NA           NA           NA NA
##  4  1973           NA           NA           NA NA
##  5  1974           NA           NA           NA NA
##  6  1975           NA           NA           NA NA
##  7  1976           NA           NA           NA NA
##  8  1977           NA           NA           NA NA
##  9  1978           NA           NA           NA NA
## 10  1979           NA           NA           NA NA
## # ... with 40 more rows, and 7 more variables: Owns_Smartphone <chr>,
## #   `15-24` <dbl>, `25-44` <dbl>, `45-64` <dbl>, `<U+F0B3>65` <dbl>,
## #   Avg_SM_Use <dbl>, Markets <dbl>
```

```r
# add in computer and home internet use

colnames(Census_compint) <- c("Year", "Comp_athome", "Internet_athome")

wide <- full_join(wide, Census_compint)
wide
```

```
## # A tibble: 50 x 14
```

```
##      Year `15-19_rate` `20-24_rate` `10-14_rate` Owns_Cellphone
##      <dbl>        <dbl>        <dbl>        <dbl> <chr>
## 1  1970           NA           NA           NA NA
## 2  1971           NA           NA           NA NA
## 3  1972           NA           NA           NA NA
## 4  1973           NA           NA           NA NA
## 5  1974           NA           NA           NA NA
## 6  1975           NA           NA           NA NA
## 7  1976           NA           NA           NA NA
## 8  1977           NA           NA           NA NA
## 9  1978           NA           NA           NA NA
## 10 1979           NA           NA           NA NA
## # ... with 40 more rows, and 9 more variables: Owns_Smartphone <chr>,
## #   `15-24` <dbl>, `25-44` <dbl>, `45-64` <dbl>, `<U+F0B3>65` <dbl>,
## #   Avg_SM_Use <dbl>, Markets <dbl>, Comp_athome <dbl>,
## #   Internet_athome <dbl>
```

```r
# creating column names

colnames(wide) <- c("Year", "15-19_rate", "20-24_rate", "10-14_rate", "Owns_Cellphone", "Owns_Smartphone


# export completed wide dataset

write.csv(wide, "wide.csv")
```