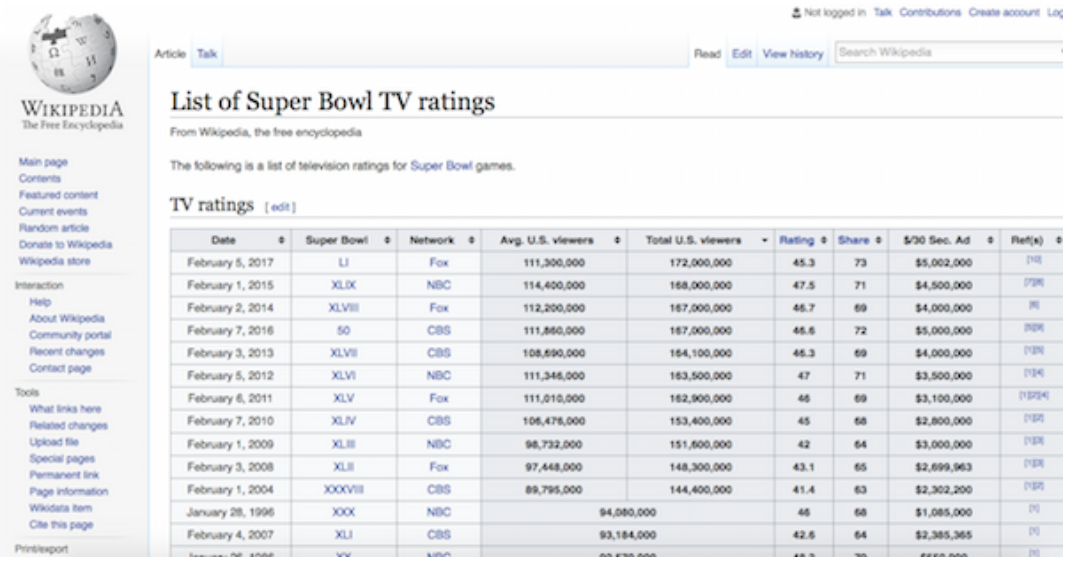
Assignment #6

Jake Moody

2/13/2017

For this assignment, I wanted to collect annual viewership data for the Superbowl. I found this data on Wikipedia. Here's a screenshot of the data as it is shown on the web:



List of Super Bowl TV ratings

From Wikipedia, the free encyclopedia

The following is a list of television ratings for Super Bowl games.

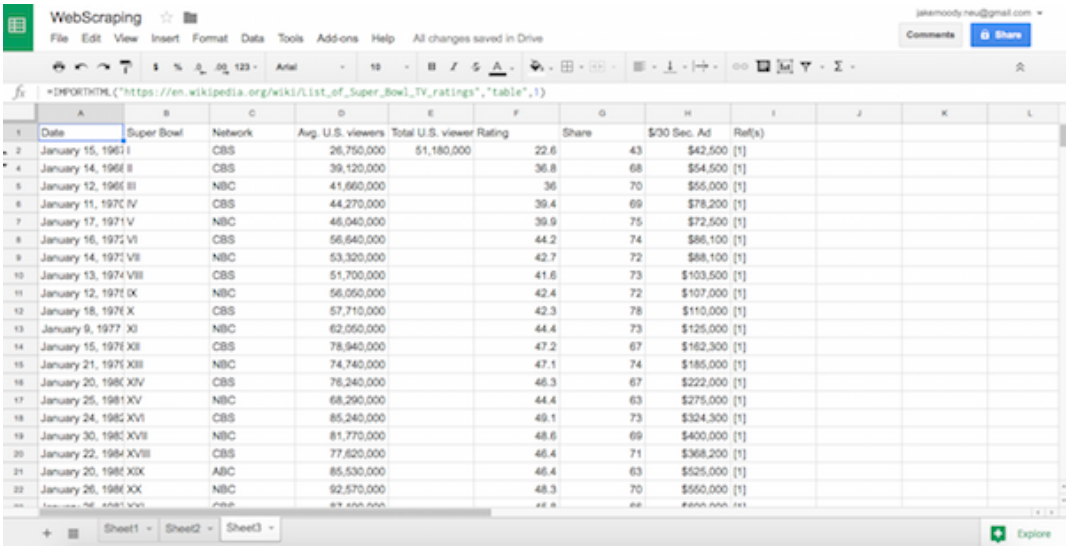
TV ratings [edit]

Date	Super Bowl	Network	Avg. U.S. viewers	Total U.S. viewers	Rating	Share	$/30 Sec. Ad	Ref(s)
February 5, 2017	LI	Fox	111,300,000	172,000,000	45.3	73	$5,002,000	
February 1, 2015	XLIX	NBC	114,400,000	168,000,000	47.5	71	$4,500,000	
February 2, 2014	XLVIII	Fox	112,200,000	167,000,000	46.7	69	$4,000,000	
February 7, 2016	50	CBS	111,860,000	167,000,000	46.6	72	$5,000,000	
February 3, 2013	XLVII	CBS	108,690,000	164,100,000	46.3	69	$4,000,000	
February 5, 2012	XLVI	NBC	111,346,000	163,500,000	47	71	$3,500,000	
February 6, 2011	XLV	Fox	111,010,000	162,900,000	46	69	$3,100,000	
February 7, 2010	XLIV	CBS	106,476,000	153,400,000	45	68	$2,800,000	
February 1, 2009	XLIII	NBC	98,732,000	151,600,000	42	64	$3,000,000	
February 3, 2008	XLII	Fox	97,448,000	148,300,000	43.1	65	$2,699,963	
February 1, 2004	XXXVIII	CBS	89,795,000	144,400,000	41.4	63	$2,302,200	
January 28, 1996	XXX	NBC	94,080,000		46	68	$1,085,000	
February 4, 2007	XLI	CBS	93,184,000		42.6	64	$2,385,365	

I used Google Sheets as a webscraping toolkit. In particular, I used the `IMPORTHTML` function which accepts a `URL`, `Query` ('Table' or 'List') and `Index` argument. Here is the function I used to scrape the Wikipedia page:

```
=IMPORTHTML("https://en.wikipedia.org/wiki/List_of_Super_Bowl_TV_ratings","table",1)
```

Here's a screenshot of the Google Sheet after I scraped the data from the Wikipedia page. You can see the function in the formula line:



WebScraping

=IMPORTHTML("https://en.wikipedia.org/wiki/List_of_Super_Bowl_TV_ratings","table",1)

	A	B	C	D	E	F	G	H	I
1	Date	Super Bowl	Network	Avg. U.S. viewers	Total U.S. viewer	Rating	Share	$/30 Sec. Ad	Ref(s)
2	January 15, 1967	I	CBS	26,750,000	51,180,000	22.6	43	$42,500	[1]
4	January 14, 1968	II	CBS	39,120,000		36.8	68	$54,500	[1]
5	January 12, 1969	III	NBC	41,660,000		36	70	$55,000	[1]
6	January 11, 1970	IV	CBS	44,270,000		39.4	69	$78,200	[1]
7	January 17, 1971	V	NBC	46,040,000		39.9	75	$72,500	[1]
8	January 16, 1972	VI	CBS	56,640,000		44.2	74	$86,100	[1]
9	January 14, 1973	VII	NBC	53,320,000		42.7	72	$88,100	[1]
10	January 13, 1974	VIII	CBS	51,700,000		41.6	73	$103,500	[1]
11	January 12, 1975	IX	NBC	56,050,000		42.4	72	$107,000	[1]
12	January 18, 1976	X	CBS	57,710,000		42.3	78	$110,000	[1]
13	January 9, 1977	XI	NBC	62,050,000		44.4	73	$125,000	[1]
14	January 15, 1978	XII	CBS	78,940,000		47.2	67	$162,300	[1]
15	January 21, 1979	XIII	NBC	74,740,000		47.1	74	$185,000	[1]
16	January 20, 1980	XIV	CBS	76,240,000		46.3	67	$222,000	[1]
17	January 25, 1981	XV	NBC	68,290,000		44.4	63	$275,000	[1]
18	January 24, 1982	XVI	CBS	85,240,000		49.1	73	$324,300	[1]
19	January 30, 1983	XVII	NBC	81,770,000		48.6	69	$400,000	[1]
20	January 22, 1984	XVIII	CBS	77,620,000		46.4	71	$368,200	[1]
21	January 20, 1985	XIX	ABC	85,530,000		46.4	63	$525,000	[1]
22	January 26, 1986	XX	NBC	92,570,000		48.3	70	$550,000	[1]

Sheet1 Sheet2 Sheet3 Explore

Once I finished scraping the data, I saved the Google Sheet as a .csv file. I then used the `read.csv` function in R to import the sheet as a data frame. You can see how I did this with the code below. I also tested

the `class` of `superbowl_views` to confirm it was a data frame. Then, I used the `print` function to show a complete output of the data frame.

```
# Import Super Bowl Viewership Data as a data frame

superbowl_views <- read.csv("WebScraping - Sheet3.csv", header = TRUE) # import

class(superbowl_views) # confirm it's a data frame
```

```
## [1] "data.frame"
```

```
print(superbowl_views) # print the results
```

```
##                Date Super.Bowl Network Avg..U.S..viewers
## 1  January 15, 1967          I     CBS        26,750,000
## 2                                 NBC        24,430,000
## 3  January 14, 1968         II     CBS        39,120,000
## 4  January 12, 1969        III     NBC        41,660,000
## 5  January 11, 1970         IV     CBS        44,270,000
## 6  January 17, 1971          V     NBC        46,040,000
## 7  January 16, 1972         VI     CBS        56,640,000
## 8  January 14, 1973        VII     NBC        53,320,000
## 9  January 13, 1974       VIII     CBS        51,700,000
## 10 January 12, 1975         IX     NBC        56,050,000
## 11 January 18, 1976          X     CBS        57,710,000
## 12  January 9, 1977         XI     NBC        62,050,000
## 13 January 15, 1978        XII     CBS        78,940,000
## 14 January 21, 1979       XIII     NBC        74,740,000
## 15 January 20, 1980        XIV     CBS        76,240,000
## 16 January 25, 1981         XV     NBC        68,290,000
## 17 January 24, 1982        XVI     CBS        85,240,000
## 18 January 30, 1983       XVII     NBC        81,770,000
## 19 January 22, 1984      XVIII     CBS        77,620,000
## 20 January 20, 1985        XIX     ABC        85,530,000
## 21 January 26, 1986         XX     NBC        92,570,000
## 22 January 25, 1987        XXI     CBS        87,190,000
## 23 January 31, 1988       XXII     ABC        80,140,000
## 24 January 22, 1989      XXIII     NBC        81,590,000
## 25 January 28, 1990       XXIV     CBS        73,852,000
## 26 January 27, 1991        XXV     ABC        79,510,000
## 27 January 26, 1992       XXVI     CBS        79,590,000
## 28 January 31, 1993      XXVII     NBC        90,990,000
## 29 January 30, 1994     XXVIII                90,000,000
## 30 January 29, 1995       XXIX     ABC        83,420,000
## 31 January 28, 1996        XXX     NBC        94,080,000
## 32 January 26, 1997       XXXI     Fox        87,870,000
## 33 January 25, 1998      XXXII     NBC        90,000,000
## 34 January 31, 1999     XXXIII     Fox        83,720,000
## 35 January 30, 2000      XXXIV     ABC        88,465,000
## 36 January 28, 2001       XXXV     CBS        84,335,000
## 37 February 3, 2002      XXXVI     Fox        86,801,000
## 38 January 26, 2003     XXXVII     ABC        88,637,000
## 39 February 1, 2004    XXXVIII     CBS        89,795,000
## 40 February 6, 2005      XXXIX     Fox        86,072,000
## 41 February 5, 2006         XL     ABC        90,745,000
```

```
## 42 February 4, 2007        XLI     CBS        93,184,000
## 43 February 3, 2008       XLII     Fox        97,448,000
## 44 February 1, 2009      XLIII     NBC        98,732,000
## 45 February 7, 2010       XLIV     CBS       106,476,000
## 46 February 6, 2011        XLV     Fox       111,010,000
## 47 February 5, 2012       XLVI     NBC       111,346,000
## 48 February 3, 2013      XLVII     CBS       108,690,000
## 49 February 2, 2014     XLVIII     Fox       112,200,000
## 50 February 1, 2015       XLIX     NBC       114,400,000
## 51 February 7, 2016         50     CBS       111,860,000
## 52 February 5, 2017         LI     Fox       111,300,000
##    Total.U.S..viewers Rating Share X..30.Sec..Ad    Ref.s.
## 1          51,180,000   22.6    43       $42,500       [1]
## 2                       18.5    36       $37,500
## 3                       36.8    68       $54,500       [1]
## 4                       36.0    70       $55,000       [1]
## 5                       39.4    69       $78,200       [1]
## 6                       39.9    75       $72,500       [1]
## 7                       44.2    74       $86,100       [1]
## 8                       42.7    72       $88,100       [1]
## 9                       41.6    73      $103,500       [1]
## 10                      42.4    72      $107,000       [1]
## 11                      42.3    78      $110,000       [1]
## 12                      44.4    73      $125,000       [1]
## 13                      47.2    67      $162,300       [1]
## 14                      47.1    74      $185,000       [1]
## 15                      46.3    67      $222,000       [1]
## 16                      44.4    63      $275,000       [1]
## 17                      49.1    73      $324,300       [1]
## 18                      48.6    69      $400,000       [1]
## 19                      46.4    71      $368,200       [1]
## 20                      46.4    63      $525,000       [1]
## 21                      48.3    70      $550,000       [1]
## 22                      45.8    66      $600,000       [1]
## 23                      41.9    62      $645,000       [1]
## 24                      43.5    68      $675,000       [1]
## 25                      39.0    67      $700,400       [1]
## 26                      41.9    63      $800,000       [1]
## 27                      40.3    61      $850,000       [1]
## 28                      45.1    66                     [1]
## 29                      45.5    66      $900,000       [1]
## 30                      41.3    62    $1,150,000       [1]
## 31                      46.0    68    $1,085,000       [1]
## 32                      43.3    65    $1,200,000       [1]
## 33                      44.5    67    $1,291,100       [1]
## 34                      40.2    61    $1,600,000       [1]
## 35                      43.3    63    $2,100,000       [1]
## 36                      40.4    61    $2,200,000       [1]
## 37                      40.4    61                     [1]
## 38                      40.7    61                     [1]
## 39        144,400,000   41.4    63    $2,302,200    [1][2]
## 40                      41.1    62    $2,400,000       [1]
## 41                      41.6    62    $2,500,000       [1]
## 42                      42.6    64    $2,385,365       [1]
```

```
## 43        148,300,000   43.1    65    $2,699,963    [1][3]
## 44        151,600,000   42.0    64    $3,000,000    [1][3]
## 45        153,400,000   45.0    68    $2,800,000    [1][2]
## 46        162,900,000   46.0    69    $3,100,000 [1][2][4]
## 47        163,500,000   47.0    71    $3,500,000    [1][4]
## 48        164,100,000   46.3    69    $4,000,000    [1][5]
## 49        167,000,000   46.7    69                     [6]
## 50        168,000,000   47.5    71    $4,500,000    [7][8]
## 51        167,000,000   46.6    72    $5,000,000    [5][9]
## 52        172,000,000   45.3    73    $5,002,000      [10]
```

The output of this data frame mirrors the data in Google Sheets and the data avaialble on Wikipedia. This data is in a interpretable format and only requires a bit of cleaning to begin analysis.