IT 497 HOMEWORK3

library(readr)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

# QUESTION 1  
# Below is a link to a file listing the number women software engineers at several Silicon Valley tech firms  
wom\_soft\_eng <- read.csv("http://www.itk.ilstu.edu/faculty/jrwolf/wit.csv", header = TRUE)

# 1 a) R code to find the number women software engineers at Reddit  
women <- filter(wom\_soft\_eng, company == "Reddit")  
  
women$num\_female\_eng

## [1] 2

# 1 b) create the R code to find the total number women software engineers in the dataset  
sum(select(wom\_soft\_eng, num\_female\_eng))

## [1] 2883

# QUESTION 2  
# The R code below, it will create a dataframe (named pop) with 139 observations and 4 variables.  
  
### Start pop Code#######  
library("rvest")

## Loading required package: xml2

##   
## Attaching package: 'rvest'

## The following object is masked from 'package:readr':  
##   
## guess\_encoding

html <- read\_html("https://www.ssa.gov/oact/babynames/numberUSbirths.html")  
usBirthData <- html\_table(html, fill=TRUE)  
  
pop <- usBirthData[[1]]  
pop$Male <- as.numeric(gsub(",", "", pop$Male))  
pop$Female <- as.numeric(gsub(",", "", pop$Female))  
pop$Total <- as.numeric(gsub(",", "", pop$Total))  
names(pop) <- c("year", "male", "female", "total")  
### End pop Code #######

# 2 a) Create the R code to find the year with the most children born.  
pop[which.max(pop$total), 'year']

## [1] 2007

# 2 b Using the same list, create the R code to find the year with the fewest children born.  
  
pop[which.min(pop$total), 'year']

## [1] 1881

#QUESTION 3 The following code will download the data from the wiki page above and create a dataframe named top Cities  
library(tidyverse)

## -- Attaching packages -------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.3 v stringr 1.4.0  
## v tidyr 1.1.2 v forcats 0.5.0

## -- Conflicts ----------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x rvest::guess\_encoding() masks readr::guess\_encoding()  
## x dplyr::lag() masks stats::lag()  
## x purrr::pluck() masks rvest::pluck()

library(dplyr)  
library("rvest")  
  
#### Start Top Cities Code ####  
topCities <- read\_html("https://en.wikipedia.org/wiki/List\_of\_cities\_proper\_by\_population")   
topCities <- html\_table( html\_node( topCities, "table.sortable" ), header=TRUE, trim=F, fill = TRUE)[-1,]   
names(topCities) <- c( "City", "Country", "Image", "Population", "City.Def", "City.Population", "City.Area", "Metro.Pop", "Metro.Area", "Urban.Pop", "Urban.Area" )   
  
topCities$Image <- NULL   
  
## Old String Substitute (gsub)   
topCities$Population <- as.numeric(gsub(",", "", topCities$Population))   
  
## tidyverse String Substitute (str\_replace\_all)   
topCities$City<-str\_replace\_all(topCities$City, "\n", "")   
topCities$Country<-str\_replace\_all(topCities$Country, "\n", "")  
topCities$City.Def<-str\_replace\_all(topCities$City.Def, "\n", "")   
topCities$City.Area<-str\_replace\_all(topCities$City.Area, "\n", "")   
topCities$Metro.Pop<-str\_replace\_all(topCities$Metro.Pop, "\n", "")   
topCities$Metro.Area<-str\_replace\_all(topCities$Metro.Area, "\n", "")   
topCities$Urban.Pop<-str\_replace\_all(topCities$Urban.Pop, "\n", "")   
topCities$Urban.Area<-str\_replace\_all(topCities$Urban.Area, "\n", "")   
#### End Top Cities Code ####

#3 a) Use the topCities dataframe and create the R code to find the 5 largest cities in India.  
topCities$Country<-str\_trim(topCities$Country)   
  
indian\_cities <- filter(topCities, Country == "India")  
  
## Rearranging rows   
indian\_cities = arrange(indian\_cities, desc(Population))  
  
head(indian\_cities, 5)

## City Country Population City.Def City.Population  
## 1 Delhi India 28514000 National capital territory 16,753,235[16]  
## 2 Mumbai India 19980000 Municipality 12,478,447[24]  
## 3 Kolkata India 14681000 Municipality 4,496,694[38]  
## 4 Bangalore India 11440000 Municipality 8,443,675[49]  
## 5 Chennai India 10456000 Municipality 6,727,000[56]  
## City.Area Metro.Pop Metro.Area Urban.Pop Urban.Area  
## 1 1,484 29,000,000[17] 3,483[17] 29,617,000 2,232[f]  
## 2 603 24,400,000[25] 4,355[26] 23,355,000 944[i]  
## 3 205 14,035,959[39] 1,851[40] 17,560,000 1,351  
## 4 709 13,707,000 1,205  
## 5 426[57] 11,324,000 1,049

#3 b. Use the topCities dataframe and create the R code to find the 5 largest cities in China.  
china\_cities <- filter(topCities, Country == "China")  
  
## Rearranging rows   
china\_cities = arrange(china\_cities, desc(Population))  
  
head(china\_cities, 5)

## City Country Population City.Def City.Population City.Area  
## 1 Shanghai China 25582000 Municipality 24,183,000[18] 6,341  
## 2 Beijing China 19618000 Municipality 21,707,000[18] 16,411  
## 3 Chongqing China 14838000 Municipality 30,165,500[36] 82,403  
## 4 Tianjin China 13215000 Municipality 15,569,000[18] 11,920  
## 5 Guangzhou China 12638000 City (sub-provincial) 14,498,400[18] 7,434  
## Metro.Pop Metro.Area Urban.Pop Urban.Area  
## 1 N/A N/A 22,120,000 4,068[g]  
## 2 N/A N/A 19,433,000 4,172  
## 3 N/A N/A 7,739,000 1,537  
## 4 N/A N/A 10,800,000 2,813  
## 5 N/A N/A 20,902,000 4,342[n]

## QUESTION 4  
## code to download the S&P 500 companies and the dates they were added.  
########## SnP Code Start ##############  
  
url <- "https://en.wikipedia.org/wiki/List\_of\_S%26P\_500\_companies"  
url %>%  
read\_html() %>%  
html\_nodes("#constituents") %>%  
html\_table(header = T) -> sp500  
sp500 <- sp500[[1]]  
names(sp500) <- c(  
"Ticker", "Name", "SEC.Filing", "GICS.Sector",  
"GICS.SubIndustry", "Location.HQ", "Date.Added", "CIK", "Founded"  
)  
sp500$Date.Added <- as.numeric(substr(sp500$Date.Added,1,4))

########## SnP Code End ##############  
# 4 a) Create the R code to find the names of firms added to the S&P 500 since 2010.  
  
# We first filter all rows of firms added since 2010  
Comp\_added <- filter(arrange(sp500, desc(Date.Added)), Date.Added >= 2010)  
Comp\_added

## Ticker Name SEC.Filing GICS.Sector  
## 1 BIO Bio-Rad Laboratories reports Health Care  
## 2 CARR Carrier Global reports Industrials  
## 3 DXCM DexCom reports Health Care  
## 4 DPZ Domino's Pizza reports Consumer Discretionary  
## 5 IR Ingersoll Rand reports Industrials  
## 6 OTIS Otis Worldwide reports Industrials  
## 7 PAYC Paycom reports Information Technology  
## 8 TDY Teledyne Technologies reports Industrials  
## 9 TYL Tyler Technologies reports Information Technology  
## 10 WST West Pharmaceutical Services reports Health Care  
## 11 AMCR Amcor plc reports Materials  
## 12 ATO Atmos Energy reports Utilities  
## 13 CDW CDW reports Information Technology  
## 14 CTVA Corteva reports Materials  
## 15 DOW Dow Inc. reports Materials  
## 16 DD DuPont de Nemours Inc reports Materials  
## 17 FRC First Republic Bank reports Financials  
## 18 IEX IDEX Corporation reports Industrials  
## 19 LVS Las Vegas Sands reports Consumer Discretionary  
## 20 LDOS Leidos Holdings reports Information Technology  
## 21 LYV Live Nation Entertainment reports Communication Services  
## 22 MKTX MarketAxess reports Financials  
## 23 NVR NVR, Inc. reports Consumer Discretionary  
## 24 ODFL Old Dominion Freight Line reports Industrials  
## 25 NOW ServiceNow reports Information Technology  
## 26 STE STERIS plc reports Health Care  
## 27 TMUS T-Mobile US reports Communication Services  
## 28 TFX Teleflex reports Health Care  
## 29 WRB W. R. Berkley Corporation reports Financials  
## 30 WAB Wabtec Corporation reports Industrials  
## 31 ZBRA Zebra Technologies reports Information Technology  
## 32 ABMD ABIOMED Inc reports Health Care  
## 33 ANET Arista Networks reports Information Technology  
## 34 BR Broadridge Financial Solutions reports Information Technology  
## 35 CE Celanese reports Materials  
## 36 CPRT Copart Inc reports Industrials  
## 37 FANG Diamondback Energy reports Energy  
## 38 EVRG Evergy reports Utilities  
## 39 FLT FleetCor Technologies Inc reports Information Technology  
## 40 FTNT Fortinet reports Information Technology  
## 41 HFC HollyFrontier Corp reports Energy  
## 42 HII Huntington Ingalls Industries reports Industrials  
## 43 IPGP IPG Photonics Corp. reports Information Technology  
## 44 JKHY Jack Henry & Associates reports Information Technology  
## 45 KEYS Keysight Technologies reports Information Technology  
## 46 LW Lamb Weston Holdings Inc reports Consumer Staples  
## 47 MXIM Maxim Integrated Products Inc reports Information Technology  
## 48 MSCI MSCI Inc reports Financials  
## 49 ROL Rollins Inc. reports Industrials  
## 50 SIVB SVB Financial reports Financials  
## 51 TTWO Take-Two Interactive reports Communication Services  
## 52 TWTR Twitter, Inc. reports Communication Services  
## 53 AMD Advanced Micro Devices Inc reports Information Technology  
## 54 ARE Alexandria Real Estate Equities reports Real Estate  
## 55 ALGN Align Technology reports Health Care  
## 56 ANSS ANSYS reports Information Technology  
## 57 AOS A.O. Smith Corp reports Industrials  
## 58 BKR Baker Hughes Co reports Energy  
## 59 CDNS Cadence Design Systems reports Information Technology  
## 60 CBOE Cboe Global Markets reports Financials  
## 61 DISH Dish Network reports Communication Services  
## 62 DRE Duke Realty Corp reports Real Estate  
## 63 DXC DXC Technology reports Information Technology  
## 64 RE Everest Re Group Ltd. reports Financials  
## 65 IT Gartner Inc reports Information Technology  
## 66 HLT Hilton Worldwide Holdings Inc reports Consumer Discretionary  
## 67 IDXX IDEXX Laboratories reports Health Care  
## 68 INFO IHS Markit Ltd. reports Industrials  
## 69 INCY Incyte reports Health Care  
## 70 IQV IQVIA Holdings Inc. reports Health Care  
## 71 MGM MGM Resorts International reports Consumer Discretionary  
## 72 NCLH Norwegian Cruise Line Holdings reports Consumer Discretionary  
## 73 PKG Packaging Corporation of America reports Materials  
## 74 RJF Raymond James Financial Inc. reports Financials  
## 75 REG Regency Centers Corporation reports Real Estate  
## 76 RMD ResMed reports Health Care  
## 77 SBAC SBA Communications reports Real Estate  
## 78 SNPS Synopsys Inc. reports Information Technology  
## 79 ALK Alaska Air Group Inc reports Industrials  
## 80 ALB Albemarle Corp reports Materials  
## 81 LNT Alliant Energy Corp reports Utilities  
## 82 AWK American Water Works Company Inc reports Utilities  
## 83 AJG Arthur J. Gallagher & Co. reports Financials  
## 84 CNC Centene Corporation reports Health Care  
## 85 CHTR Charter Communications reports Communication Services  
## 86 CFG Citizens Financial Group reports Financials  
## 87 CXO Concho Resources reports Energy  
## 88 COO The Cooper Companies reports Health Care  
## 89 COTY Coty, Inc reports Consumer Staples  
## 90 DLR Digital Realty Trust Inc reports Real Estate  
## 91 EXR Extra Space Storage reports Real Estate  
## 92 FRT Federal Realty Investment Trust reports Real Estate  
## 93 FTV Fortive Corp reports Industrials  
## 94 FBHS Fortune Brands Home & Security reports Industrials  
## 95 GPN Global Payments Inc. reports Information Technology  
## 96 HOLX Hologic reports Health Care  
## 97 LKQ LKQ Corporation reports Consumer Discretionary  
## 98 MTD Mettler Toledo reports Health Care  
## 99 MAA Mid-America Apartments reports Real Estate  
## 100 TDG TransDigm Group reports Industrials  
## 101 UDR UDR, Inc. reports Real Estate  
## 102 ULTA Ulta Beauty reports Consumer Discretionary  
## 103 UA Under Armour (Class C) reports Consumer Discretionary  
## 104 WLTW Willis Towers Watson reports Financials  
## 105 ATVI Activision Blizzard reports Communication Services  
## 106 AAP Advance Auto Parts reports Consumer Discretionary  
## 107 AAL American Airlines Group reports Industrials  
## 108 CHD Church & Dwight reports Consumer Staples  
## 109 CMCSA Comcast Corp. reports Communication Services  
## 110 EQIX Equinix reports Real Estate  
## 111 FOX Fox Corporation (Class B) reports Communication Services  
## 112 HBI Hanesbrands Inc reports Consumer Discretionary  
## 113 HCA HCA Healthcare reports Health Care  
## 114 HSIC Henry Schein reports Health Care  
## 115 HPE Hewlett Packard Enterprise reports Information Technology  
## 116 ILMN Illumina Inc reports Health Care  
## 117 JBHT J. B. Hunt Transport Services reports Industrials  
## 118 KHC Kraft Heinz Co reports Consumer Staples  
## 119 NWS News Corp. Class B reports Communication Services  
## 120 PYPL PayPal reports Information Technology  
## 121 QRVO Qorvo reports Information Technology  
## 122 O Realty Income Corporation reports Real Estate  
## 123 SWKS Skyworks Solutions reports Information Technology  
## 124 SLG SL Green Realty reports Real Estate  
## 125 SYF Synchrony Financial reports Financials  
## 126 UAL United Airlines Holdings reports Industrials  
## 127 VRSK Verisk Analytics reports Industrials  
## 128 GOOGL Alphabet Inc. (Class A) reports Communication Services  
## 129 AVGO Broadcom Inc. reports Information Technology  
## 130 DISCK Discovery, Inc. (Class C) reports Communication Services  
## 131 ESS Essex Property Trust, Inc. reports Real Estate  
## 132 MLM Martin Marietta Materials reports Materials  
## 133 RCL Royal Caribbean Group reports Consumer Discretionary  
## 134 TSCO Tractor Supply Company reports Consumer Discretionary  
## 135 UAA Under Armour (Class A) reports Consumer Discretionary  
## 136 URI United Rentals, Inc. reports Industrials  
## 137 UHS Universal Health Services reports Health Care  
## 138 ALLE Allegion reports Industrials  
## 139 AME AMETEK Inc. reports Industrials  
## 140 DAL Delta Air Lines Inc. reports Industrials  
## 141 FB Facebook, Inc. reports Communication Services  
## 142 FOXA Fox Corporation (Class A) reports Communication Services  
## 143 GM General Motors reports Consumer Discretionary  
## 144 KSU Kansas City Southern reports Industrials  
## 145 MHK Mohawk Industries reports Consumer Discretionary  
## 146 NWSA News Corp. Class A reports Communication Services  
## 147 NLSN Nielsen Holdings reports Industrials  
## 148 PVH PVH Corp. reports Consumer Discretionary  
## 149 REGN Regeneron Pharmaceuticals reports Health Care  
## 150 VRTX Vertex Pharmaceuticals Inc reports Health Care  
## 151 ZTS Zoetis reports Health Care  
## 152 ABBV AbbVie Inc. reports Health Care  
## 153 ALXN Alexion Pharmaceuticals reports Health Care  
## 154 APTV Aptiv PLC reports Consumer Discretionary  
## 155 CCI Crown Castle International Corp. reports Real Estate  
## 156 DG Dollar General reports Consumer Discretionary  
## 157 GRMN Garmin Ltd. reports Consumer Discretionary  
## 158 KMI Kinder Morgan reports Energy  
## 159 LRCX Lam Research reports Information Technology  
## 160 LYB LyondellBasell reports Materials  
## 161 MDLZ Mondelez International reports Consumer Staples  
## 162 MNST Monster Beverage reports Consumer Staples  
## 163 PNR Pentair plc reports Industrials  
## 164 PSX Phillips 66 reports Energy  
## 165 STX Seagate Technology reports Information Technology  
## 166 ACN Accenture plc reports Information Technology  
## 167 BLK BlackRock reports Financials  
## 168 BWA BorgWarner reports Consumer Discretionary  
## 169 CMG Chipotle Mexican Grill reports Consumer Discretionary  
## 170 DLTR Dollar Tree reports Consumer Discretionary  
## 171 EW Edwards Lifesciences reports Health Care  
## 172 MPC Marathon Petroleum reports Energy  
## 173 MOS The Mosaic Company reports Materials  
## 174 PRGO Perrigo reports Health Care  
## 175 TEL TE Connectivity Ltd. reports Information Technology  
## 176 XYL Xylem Inc. reports Industrials  
## 177 BRK.B Berkshire Hathaway reports Financials  
## 178 KMX Carmax Inc reports Consumer Discretionary  
## 179 CERN Cerner reports Health Care  
## 180 CB Chubb Limited reports Financials  
## 181 DISCA Discovery, Inc. (Class A) reports Communication Services  
## 182 FFIV F5 Networks reports Information Technology  
## 183 JCI Johnson Controls International reports Industrials  
## 184 NFLX Netflix Inc. reports Communication Services  
## 185 NRG NRG Energy reports Utilities  
## 186 OKE ONEOK reports Energy  
## 187 TT Trane Technologies plc reports Industrials  
## GICS.SubIndustry  
## 1 Life Sciences Tools & Services  
## 2 Building Products  
## 3 Health Care Equipment  
## 4 Restaurants  
## 5 Industrial Machinery  
## 6 Industrial Machinery  
## 7 Application Software  
## 8 Aerospace & Defense  
## 9 Application Software  
## 10 Health Care Supplies  
## 11 Paper Packaging  
## 12 Gas Utilities  
## 13 Technology Distributors  
## 14 Fertilizers & Agricultural Chemicals  
## 15 Commodity Chemicals  
## 16 Specialty Chemicals  
## 17 Regional Banks  
## 18 Industrial Machinery  
## 19 Casinos & Gaming  
## 20 IT Consulting & Other Services  
## 21 Movies & Entertainment  
## 22 Financial Exchanges & Data  
## 23 Homebuilding  
## 24 Trucking  
## 25 Systems Software  
## 26 Health Care Equipment  
## 27 Wireless Telecommunication Services  
## 28 Health Care Equipment  
## 29 Property & Casualty Insurance  
## 30 Construction Machinery & Heavy Trucks  
## 31 Electronic Equipment & Instruments  
## 32 Health Care Equipment  
## 33 Communications Equipment  
## 34 Data Processing & Outsourced Services  
## 35 Specialty Chemicals  
## 36 Diversified Support Services  
## 37 Oil & Gas Exploration & Production  
## 38 Electric Utilities  
## 39 Data Processing & Outsourced Services  
## 40 Systems Software  
## 41 Oil & Gas Refining & Marketing  
## 42 Aerospace & Defense  
## 43 Electronic Manufacturing Services  
## 44 Data Processing & Outsourced Services  
## 45 Electronic Equipment & Instruments  
## 46 Packaged Foods & Meats  
## 47 Semiconductors  
## 48 Financial Exchanges & Data  
## 49 Environmental & Facilities Services  
## 50 Regional Banks  
## 51 Interactive Home Entertainment  
## 52 Interactive Media & Services  
## 53 Semiconductors  
## 54 Office REITs  
## 55 Health Care Supplies  
## 56 Application Software  
## 57 Building Products  
## 58 Oil & Gas Equipment & Services  
## 59 Application Software  
## 60 Financial Exchanges & Data  
## 61 Cable & Satellite  
## 62 Industrial REITs  
## 63 IT Consulting & Other Services  
## 64 Reinsurance  
## 65 IT Consulting & Other Services  
## 66 Hotels, Resorts & Cruise Lines  
## 67 Health Care Equipment  
## 68 Research & Consulting Services  
## 69 Biotechnology  
## 70 Life Sciences Tools & Services  
## 71 Casinos & Gaming  
## 72 Hotels, Resorts & Cruise Lines  
## 73 Paper Packaging  
## 74 Investment Banking & Brokerage  
## 75 Retail REITs  
## 76 Health Care Equipment  
## 77 Specialized REITs  
## 78 Application Software  
## 79 Airlines  
## 80 Specialty Chemicals  
## 81 Electric Utilities  
## 82 Water Utilities  
## 83 Insurance Brokers  
## 84 Managed Health Care  
## 85 Cable & Satellite  
## 86 Regional Banks  
## 87 Oil & Gas Exploration & Production  
## 88 Health Care Supplies  
## 89 Personal Products  
## 90 Specialized REITs  
## 91 Specialized REITs  
## 92 Retail REITs  
## 93 Industrial Machinery  
## 94 Building Products  
## 95 Data Processing & Outsourced Services  
## 96 Health Care Equipment  
## 97 Distributors  
## 98 Life Sciences Tools & Services  
## 99 Residential REITs  
## 100 Aerospace & Defense  
## 101 Residential REITs  
## 102 Specialty Stores  
## 103 Apparel, Accessories & Luxury Goods  
## 104 Insurance Brokers  
## 105 Interactive Home Entertainment  
## 106 Automotive Retail  
## 107 Airlines  
## 108 Household Products  
## 109 Cable & Satellite  
## 110 Specialized REITs  
## 111 Movies & Entertainment  
## 112 Apparel, Accessories & Luxury Goods  
## 113 Health Care Facilities  
## 114 Health Care Distributors  
## 115 Technology Hardware, Storage & Peripherals  
## 116 Life Sciences Tools & Services  
## 117 Trucking  
## 118 Packaged Foods & Meats  
## 119 Publishing  
## 120 Data Processing & Outsourced Services  
## 121 Semiconductors  
## 122 Retail REITs  
## 123 Semiconductors  
## 124 Office REITs  
## 125 Consumer Finance  
## 126 Airlines  
## 127 Research & Consulting Services  
## 128 Interactive Media & Services  
## 129 Semiconductors  
## 130 Broadcasting  
## 131 Residential REITs  
## 132 Construction Materials  
## 133 Hotels, Resorts & Cruise Lines  
## 134 Specialty Stores  
## 135 Apparel, Accessories & Luxury Goods  
## 136 Trading Companies & Distributors  
## 137 Health Care Facilities  
## 138 Building Products  
## 139 Electrical Components & Equipment  
## 140 Airlines  
## 141 Interactive Media & Services  
## 142 Movies & Entertainment  
## 143 Automobile Manufacturers  
## 144 Railroads  
## 145 Home Furnishings  
## 146 Publishing  
## 147 Research & Consulting Services  
## 148 Apparel, Accessories & Luxury Goods  
## 149 Biotechnology  
## 150 Biotechnology  
## 151 Pharmaceuticals  
## 152 Pharmaceuticals  
## 153 Biotechnology  
## 154 Auto Parts & Equipment  
## 155 Specialized REITs  
## 156 General Merchandise Stores  
## 157 Consumer Electronics  
## 158 Oil & Gas Storage & Transportation  
## 159 Semiconductor Equipment  
## 160 Specialty Chemicals  
## 161 Packaged Foods & Meats  
## 162 Soft Drinks  
## 163 Industrial Machinery  
## 164 Oil & Gas Refining & Marketing  
## 165 Technology Hardware, Storage & Peripherals  
## 166 IT Consulting & Other Services  
## 167 Asset Management & Custody Banks  
## 168 Auto Parts & Equipment  
## 169 Restaurants  
## 170 General Merchandise Stores  
## 171 Health Care Equipment  
## 172 Oil & Gas Refining & Marketing  
## 173 Fertilizers & Agricultural Chemicals  
## 174 Pharmaceuticals  
## 175 Electronic Manufacturing Services  
## 176 Industrial Machinery  
## 177 Multi-Sector Holdings  
## 178 Specialty Stores  
## 179 Health Care Technology  
## 180 Property & Casualty Insurance  
## 181 Broadcasting  
## 182 Communications Equipment  
## 183 Building Products  
## 184 Movies & Entertainment  
## 185 Independent Power Producers & Energy Traders  
## 186 Oil & Gas Storage & Transportation  
## 187 Building Products  
## Location.HQ Date.Added CIK  
## 1 Hercules, California 2020 12208  
## 2 Palm Beach Gardens, Florida 2020 1783180  
## 3 San Diego, California 2020 1093557  
## 4 Ann Arbor, Michigan 2020 1286681  
## 5 Milwaukee, Wisconsin 2020 1699150  
## 6 Farmington, Connecticut 2020 1781335  
## 7 Oklahoma City, Oklahoma 2020 1590955  
## 8 Thousand Oaks, California 2020 1094285  
## 9 Plano, Texas 2020 860731  
## 10 Exton, Pennsylvania 2020 105770  
## 11 Warmley, Bristol, United Kingdom 2019 1748790  
## 12 Dallas, Texas 2019 731802  
## 13 Lincolnshire, Illinois 2019 1402057  
## 14 Wilmington, Delaware 2019 1755672  
## 15 Midland, Michigan 2019 1751788  
## 16 Midland, Michigan 2019 1666700  
## 17 San Francisco, California 2019 1132979  
## 18 Lake Forest, Illinois 2019 832101  
## 19 Las Vegas, Nevada 2019 1300514  
## 20 Reston, Virginia 2019 1336920  
## 21 Beverly Hills, California 2019 1335258  
## 22 New York, New York 2019 1278021  
## 23 Reston, Virginia 2019 906163  
## 24 Thomasville, North Carolina 2019 878927  
## 25 Santa Clara, California 2019 1373715  
## 26 Dublin, Ireland 2019 1757898  
## 27 Bellevue, Washington 2019 1283699  
## 28 Wayne, Pennsylvania 2019 96943  
## 29 Greenwich, Connecticut 2019 11544  
## 30 Wilmerding, Pennsylvania 2019 943452  
## 31 Lincolnshire, Illinois 2019 877212  
## 32 Danvers, Massachusetts 2018 815094  
## 33 Santa Clara, California 2018 1596532  
## 34 Lake Success, New York 2018 1383312  
## 35 Irving, Texas 2018 1306830  
## 36 Dallas, Texas 2018 900075  
## 37 Midland, Texas 2018 1539838  
## 38 Kansas City, Missouri 2018 1711269  
## 39 Norcross, Georgia 2018 1175454  
## 40 Sunnyvale, California 2018 1262039  
## 41 Dallas, Texas 2018 48039  
## 42 Newport News, Virginia 2018 1501585  
## 43 Oxford, Massachusetts 2018 1111928  
## 44 Monett, Missouri 2018 779152  
## 45 Santa Rosa, California 2018 1601046  
## 46 Eagle, Idaho 2018 1679273  
## 47 San Jose, California 2018 743316  
## 48 New York, New York 2018 1408198  
## 49 Atlanta, Georgia 2018 84839  
## 50 Santa Clara, California 2018 719739  
## 51 New York, New York 2018 946581  
## 52 San Francisco, California 2018 1418091  
## 53 Santa Clara, California 2017 2488  
## 54 Pasadena, California 2017 1035443  
## 55 San Jose, California 2017 1097149  
## 56 Canonsburg, Pennsylvania 2017 1013462  
## 57 Milwaukee, Wisconsin 2017 91142  
## 58 Houston, Texas 2017 1701605  
## 59 San Jose, California 2017 813672  
## 60 Chicago, Illinois 2017 1374310  
## 61 Meridian, Colorado 2017 1001082  
## 62 Indianapolis, Indiana 2017 783280  
## 63 Tysons Corner, Virginia 2017 1688568  
## 64 Hamilton, Bermuda 2017 1095073  
## 65 Stamford, Connecticut 2017 749251  
## 66 Tysons Corner, Virginia 2017 1585689  
## 67 Westbrook, Maine 2017 874716  
## 68 London, United Kingdom 2017 1598014  
## 69 Wilmington, Delaware 2017 879169  
## 70 Durham, North Carolina 2017 1478242  
## 71 Paradise, Nevada 2017 789570  
## 72 Miami, Florida 2017 1513761  
## 73 Lake Forest, Illinois 2017 75677  
## 74 St. Petersburg, Florida 2017 720005  
## 75 Jacksonville, Florida 2017 910606  
## 76 San Diego, California 2017 943819  
## 77 Boca Raton, Florida 2017 1034054  
## 78 Mountain View, California 2017 883241  
## 79 Seattle, Washington 2016 766421  
## 80 Charlotte, North Carolina 2016 915913  
## 81 Madison, Wisconsin 2016 352541  
## 82 Camden, New Jersey 2016 1410636  
## 83 Itasca, Illinois 2016 354190  
## 84 St. Louis, Missouri 2016 1071739  
## 85 Stamford, Connecticut 2016 1091667  
## 86 Providence, Rhode Island 2016 759944  
## 87 Midland, Texas 2016 1358071  
## 88 San Ramon, California 2016 711404  
## 89 New York, New York 2016 1024305  
## 90 San Francisco, California 2016 1297996  
## 91 Salt Lake City, Utah 2016 1289490  
## 92 Rockville, Maryland 2016 34903  
## 93 Everett, Washington 2016 1659166  
## 94 Deerfield, Illinois 2016 1519751  
## 95 Atlanta, Georgia 2016 1123360  
## 96 Marlborough, Massachusetts 2016 859737  
## 97 Chicago, Illinois 2016 1065696  
## 98 Columbus, Ohio 2016 1037646  
## 99 Memphis, Tennessee 2016 912595  
## 100 Cleveland, Ohio 2016 1260221  
## 101 Highlands Ranch, Colorado 2016 74208  
## 102 Bolingbrook, Illinois 2016 1403568  
## 103 Baltimore, Maryland 2016 1336917  
## 104 London, United Kingdom 2016 1140536  
## 105 Santa Monica, California 2015 718877  
## 106 Raleigh, North Carolina 2015 1158449  
## 107 Fort Worth, Texas 2015 6201  
## 108 Ewing, New Jersey 2015 313927  
## 109 Philadelphia, Pennsylvania 2015 1166691  
## 110 Redwood City, California 2015 1101239  
## 111 New York, New York 2015 1308161  
## 112 Winston-Salem, North Carolina 2015 1359841  
## 113 Nashville, Tennessee 2015 860730  
## 114 Melville, New York 2015 1000228  
## 115 Palo Alto, California 2015 1645590  
## 116 San Diego, California 2015 1110803  
## 117 Lowell, Arkansas 2015 728535  
## 118 Chicago, Illinois; Pittsburgh, Pennsylvania 2015 1637459  
## 119 New York, New York 2015 1564708  
## 120 San Jose, California 2015 1633917  
## 121 Greensboro, North Carolina 2015 1604778  
## 122 San Diego, California 2015 726728  
## 123 Woburn, Massachusetts 2015 4127  
## 124 New York, New York 2015 1040971  
## 125 Stamford, Connecticut 2015 1601712  
## 126 Chicago, Illinois 2015 100517  
## 127 Jersey City, New Jersey 2015 1442145  
## 128 Mountain View, California 2014 1652044  
## 129 San Jose, California 2014 1730168  
## 130 Silver Spring, Maryland 2014 1437107  
## 131 San Mateo, California 2014 920522  
## 132 Raleigh, North Carolina 2014 916076  
## 133 Miami, Florida 2014 884887  
## 134 Brentwood, Tennessee 2014 916365  
## 135 Baltimore, Maryland 2014 1336917  
## 136 Stamford, Connecticut 2014 1067701  
## 137 King of Prussia, Pennsylvania 2014 352915  
## 138 Dublin, Ireland 2013 1579241  
## 139 Berwyn, Pennsylvania 2013 1037868  
## 140 Atlanta, Georgia 2013 27904  
## 141 Menlo Park, California 2013 1326801  
## 142 New York, New York 2013 1308161  
## 143 Detroit, Michigan 2013 1467858  
## 144 Kansas City, Missouri 2013 54480  
## 145 Calhoun, Georgia 2013 851968  
## 146 New York, New York 2013 1564708  
## 147 New York, New York 2013 1492633  
## 148 New York, New York 2013 78239  
## 149 Tarrytown, New York 2013 872589  
## 150 Cambridge, Massachusetts 2013 875320  
## 151 Florham Park, New Jersey 2013 1555280  
## 152 North Chicago, Illinois 2012 1551152  
## 153 Boston, Massachusetts 2012 899866  
## 154 Dublin, Ireland 2012 1521332  
## 155 Houston, Texas 2012 1051470  
## 156 Goodlettsville, Tennessee 2012 29534  
## 157 Schaffhausen, Switzerland 2012 1121788  
## 158 Houston, Texas 2012 1506307  
## 159 Fremont, California 2012 707549  
## 160 Rotterdam, Netherlands 2012 1489393  
## 161 Chicago, Illinois 2012 1103982  
## 162 Corona, California 2012 865752  
## 163 Worsley, UK 2012 77360  
## 164 Houston, Texas 2012 1534701  
## 165 Dublin, Ireland 2012 1137789  
## 166 Dublin, Ireland 2011 1467373  
## 167 New York, New York 2011 1364742  
## 168 Auburn Hills, Michigan 2011 908255  
## 169 Newport Beach, California 2011 1058090  
## 170 Chesapeake, Virginia 2011 935703  
## 171 Irvine, California 2011 1099800  
## 172 Findlay, Ohio 2011 1510295  
## 173 Tampa, Florida 2011 1285785  
## 174 Dublin, Ireland 2011 1585364  
## 175 Schaffhausen, Switzerland 2011 1385157  
## 176 White Plains, New York 2011 1524472  
## 177 Omaha, Nebraska 2010 1067983  
## 178 Richmond, Virginia 2010 1170010  
## 179 North Kansas City, Missouri 2010 804753  
## 180 Zurich, Switzerland 2010 896159  
## 181 Silver Spring, Maryland 2010 1437107  
## 182 Seattle, Washington 2010 1048695  
## 183 Cork, Ireland 2010 833444  
## 184 Los Gatos, California 2010 1065280  
## 185 Princeton, New Jersey 2010 1013871  
## 186 Tulsa, Oklahoma 2010 1039684  
## 187 Dublin, Ireland 2010 1466258  
## Founded  
## 1 1952  
## 2 2020 (1915)  
## 3 1999  
## 4 1960  
## 5 1859  
## 6 2020 (1853)  
## 7 1998  
## 8 1960  
## 9 1966  
## 10 1923  
## 11 2019 (1860)  
## 12 1906  
## 13 1984  
## 14 2019  
## 15 2019  
## 16 2017  
## 17 1985  
## 18 1988  
## 19 1988  
## 20 1969  
## 21 2010  
## 22 2000  
## 23 1980  
## 24 1934  
## 25 2003  
## 26 1985  
## 27 1994  
## 28 1943  
## 29 1967  
## 30 1999 (1869)  
## 31 1969  
## 32 1981  
## 33 2004  
## 34 1962  
## 35 1918  
## 36 1982  
## 37 2007  
## 38 1909  
## 39 2000  
## 40 2000  
## 41 1947  
## 42 2011  
## 43 1990  
## 44 1976  
## 45 2014  
## 46 2016 (1950)  
## 47 1983  
## 48 1969  
## 49 1948  
## 50 1983  
## 51 1993  
## 52 2006  
## 53 1969  
## 54   
## 55 1997  
## 56 1969  
## 57 1916  
## 58 2017  
## 59 1988  
## 60 1973  
## 61 1980  
## 62 1972  
## 63 2017  
## 64 1973  
## 65 1979  
## 66 1919  
## 67 1983  
## 68 1959  
## 69 1991  
## 70 1982  
## 71 1986  
## 72 2011 (1966)  
## 73 1959  
## 74 1962  
## 75 1963  
## 76 1989  
## 77   
## 78 1986  
## 79 1985  
## 80 1994  
## 81 1917  
## 82 1886  
## 83 1927  
## 84 1984  
## 85 1993  
## 86 1828  
## 87 2004  
## 88 1958  
## 89 1904  
## 90 2004  
## 91 1977  
## 92 1962  
## 93 2016  
## 94 2011 (1969)  
## 95 2000  
## 96 1985  
## 97 1998  
## 98 1945  
## 99 1977  
## 100 1993  
## 101 1972  
## 102 1990  
## 103 1996  
## 104 2016  
## 105 2008  
## 106 1932  
## 107 1934  
## 108 1847  
## 109 1963  
## 110 1998  
## 111   
## 112 2000  
## 113 1968  
## 114 1932  
## 115 2015  
## 116 1998  
## 117 1961  
## 118 2015 (1869)  
## 119 2013 (1980)  
## 120 1998  
## 121 2015  
## 122 1969  
## 123 2002  
## 124 1997  
## 125 2003  
## 126 1967  
## 127 1971  
## 128 1998  
## 129 1961  
## 130 1985  
## 131 1971  
## 132 1993  
## 133 1997  
## 134 1938  
## 135 1996  
## 136 1997  
## 137 1979  
## 138 1908  
## 139 1930  
## 140 1929  
## 141 2004  
## 142   
## 143 1908  
## 144 1887  
## 145 1878  
## 146 2013 (1980)  
## 147 1923  
## 148 1881  
## 149 1988  
## 150 1989  
## 151 1952  
## 152 2013 (1888)  
## 153 1992  
## 154 1994  
## 155 1994  
## 156 1939  
## 157 1989  
## 158 1997  
## 159 1980  
## 160 2007  
## 161 2012  
## 162 2012 (1935)  
## 163 1966  
## 164 2012 (1917)  
## 165 1979  
## 166 1989  
## 167 1988  
## 168 1880  
## 169 1993  
## 170 1986  
## 171 1958  
## 172 2009 (1887)  
## 173 2004 (1865 / 1909)  
## 174 2013 (1887)  
## 175 2007  
## 176 2011  
## 177 1839  
## 178 1993  
## 179 1979  
## 180 1985  
## 181 1985  
## 182 1996  
## 183 1885  
## 184 1997  
## 185 1992  
## 186 1906  
## 187 1871

# We display the names of firms added since 2010  
Comp\_added$Name

## [1] "Bio-Rad Laboratories" "Carrier Global"   
## [3] "DexCom" "Domino's Pizza"   
## [5] "Ingersoll Rand" "Otis Worldwide"   
## [7] "Paycom" "Teledyne Technologies"   
## [9] "Tyler Technologies" "West Pharmaceutical Services"   
## [11] "Amcor plc" "Atmos Energy"   
## [13] "CDW" "Corteva"   
## [15] "Dow Inc." "DuPont de Nemours Inc"   
## [17] "First Republic Bank" "IDEX Corporation"   
## [19] "Las Vegas Sands" "Leidos Holdings"   
## [21] "Live Nation Entertainment" "MarketAxess"   
## [23] "NVR, Inc." "Old Dominion Freight Line"   
## [25] "ServiceNow" "STERIS plc"   
## [27] "T-Mobile US" "Teleflex"   
## [29] "W. R. Berkley Corporation" "Wabtec Corporation"   
## [31] "Zebra Technologies" "ABIOMED Inc"   
## [33] "Arista Networks" "Broadridge Financial Solutions"   
## [35] "Celanese" "Copart Inc"   
## [37] "Diamondback Energy" "Evergy"   
## [39] "FleetCor Technologies Inc" "Fortinet"   
## [41] "HollyFrontier Corp" "Huntington Ingalls Industries"   
## [43] "IPG Photonics Corp." "Jack Henry & Associates"   
## [45] "Keysight Technologies" "Lamb Weston Holdings Inc"   
## [47] "Maxim Integrated Products Inc" "MSCI Inc"   
## [49] "Rollins Inc." "SVB Financial"   
## [51] "Take-Two Interactive" "Twitter, Inc."   
## [53] "Advanced Micro Devices Inc" "Alexandria Real Estate Equities"   
## [55] "Align Technology" "ANSYS"   
## [57] "A.O. Smith Corp" "Baker Hughes Co"   
## [59] "Cadence Design Systems" "Cboe Global Markets"   
## [61] "Dish Network" "Duke Realty Corp"   
## [63] "DXC Technology" "Everest Re Group Ltd."   
## [65] "Gartner Inc" "Hilton Worldwide Holdings Inc"   
## [67] "IDEXX Laboratories" "IHS Markit Ltd."   
## [69] "Incyte" "IQVIA Holdings Inc."   
## [71] "MGM Resorts International" "Norwegian Cruise Line Holdings"   
## [73] "Packaging Corporation of America" "Raymond James Financial Inc."   
## [75] "Regency Centers Corporation" "ResMed"   
## [77] "SBA Communications" "Synopsys Inc."   
## [79] "Alaska Air Group Inc" "Albemarle Corp"   
## [81] "Alliant Energy Corp" "American Water Works Company Inc"  
## [83] "Arthur J. Gallagher & Co." "Centene Corporation"   
## [85] "Charter Communications" "Citizens Financial Group"   
## [87] "Concho Resources" "The Cooper Companies"   
## [89] "Coty, Inc" "Digital Realty Trust Inc"   
## [91] "Extra Space Storage" "Federal Realty Investment Trust"   
## [93] "Fortive Corp" "Fortune Brands Home & Security"   
## [95] "Global Payments Inc." "Hologic"   
## [97] "LKQ Corporation" "Mettler Toledo"   
## [99] "Mid-America Apartments" "TransDigm Group"   
## [101] "UDR, Inc." "Ulta Beauty"   
## [103] "Under Armour (Class C)" "Willis Towers Watson"   
## [105] "Activision Blizzard" "Advance Auto Parts"   
## [107] "American Airlines Group" "Church & Dwight"   
## [109] "Comcast Corp." "Equinix"   
## [111] "Fox Corporation (Class B)" "Hanesbrands Inc"   
## [113] "HCA Healthcare" "Henry Schein"   
## [115] "Hewlett Packard Enterprise" "Illumina Inc"   
## [117] "J. B. Hunt Transport Services" "Kraft Heinz Co"   
## [119] "News Corp. Class B" "PayPal"   
## [121] "Qorvo" "Realty Income Corporation"   
## [123] "Skyworks Solutions" "SL Green Realty"   
## [125] "Synchrony Financial" "United Airlines Holdings"   
## [127] "Verisk Analytics" "Alphabet Inc. (Class A)"   
## [129] "Broadcom Inc." "Discovery, Inc. (Class C)"   
## [131] "Essex Property Trust, Inc." "Martin Marietta Materials"   
## [133] "Royal Caribbean Group" "Tractor Supply Company"   
## [135] "Under Armour (Class A)" "United Rentals, Inc."   
## [137] "Universal Health Services" "Allegion"   
## [139] "AMETEK Inc." "Delta Air Lines Inc."   
## [141] "Facebook, Inc." "Fox Corporation (Class A)"   
## [143] "General Motors" "Kansas City Southern"   
## [145] "Mohawk Industries" "News Corp. Class A"   
## [147] "Nielsen Holdings" "PVH Corp."   
## [149] "Regeneron Pharmaceuticals" "Vertex Pharmaceuticals Inc"   
## [151] "Zoetis" "AbbVie Inc."   
## [153] "Alexion Pharmaceuticals" "Aptiv PLC"   
## [155] "Crown Castle International Corp." "Dollar General"   
## [157] "Garmin Ltd." "Kinder Morgan"   
## [159] "Lam Research" "LyondellBasell"   
## [161] "Mondelez International" "Monster Beverage"   
## [163] "Pentair plc" "Phillips 66"   
## [165] "Seagate Technology" "Accenture plc"   
## [167] "BlackRock" "BorgWarner"   
## [169] "Chipotle Mexican Grill" "Dollar Tree"   
## [171] "Edwards Lifesciences" "Marathon Petroleum"   
## [173] "The Mosaic Company" "Perrigo"   
## [175] "TE Connectivity Ltd." "Xylem Inc."   
## [177] "Berkshire Hathaway" "Carmax Inc"   
## [179] "Cerner" "Chubb Limited"   
## [181] "Discovery, Inc. (Class A)" "F5 Networks"   
## [183] "Johnson Controls International" "Netflix Inc."   
## [185] "NRG Energy" "ONEOK"   
## [187] "Trane Technologies plc"

# 4 b) Create the R code to find the total number of firms in the table.  
nrow(select(sp500, Name))

## [1] 505

# QUESTION 5  
# The code below will download the data from Wikipedia and create a dataframe named venue.  
### Start Venue R Code ####  
  
cricket\_world\_cup <- read\_html("http://en.wikipedia.org/wiki/2015\_Cricket\_World\_Cup")  
venues = html\_table(html\_nodes(cricket\_world\_cup, "table")[[5]])  
  
### End Venue R Code ####

# 5a) Use the venue dataframe to find the name of the venue with the largest capacity.  
  
## We transform the data capacity column to numeric.  
venues$Capacity <-gsub(",","",venues$Capacity)  
venues$Capacity <-as.numeric(venues$Capacity)  
venues[which.max(venues$Capacity), 'Venue']

## [1] "Melbourne Cricket Ground"

#5b) Use the venue dataframe to find the name of the venue with the smallest capacity  
venues[which.min(venues$Capacity), 'Venue']

## [1] "Saxton Oval"