Baseball Data

Eric Breton

April 15, 2019

```
install.packages("XML")
```

```
## Installing package into 'C:/Users/ebret/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
install.packages("RCurl")
```

```
## Installing package into 'C:/Users/ebret/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
install.packages("RVest")
```

```
## Installing package into 'C:/Users/ebret/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
install.packages("dplyr")
```

```
## Installing package into 'C:/Users/ebret/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
install.packages("reshape2")
```

```
## Installing package into 'C:/Users/ebret/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
install.packages("tidyr")
```

```
## Installing package into 'C:/Users/ebret/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
install.packages("data.table")
```

```
## Installing package into 'C:/Users/ebret/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)
```

```
## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
install.packages("stringi")
```

```
## Installing package into 'C:/Users/ebret/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)
```

```
## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
library(XML)
```

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

```
library(RCurl)
```

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

```
## Loading required package: bitops
```

```
library(rvest)
```

```
## Loading required package: xml2
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:XML':
##
##     xml
```

```
library(dplyr)
```

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

```
##
## 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
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.5.3
```

```
library(tidyr)
```

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

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:reshape2':
##
##     smiths
```

```
## The following object is masked from 'package:RCurl':
##
##     complete
```

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.5.3
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:reshape2':
##
##     dcast, melt
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library(stringi)
```

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

2018

```
url<-"https://www.baseball-reference.com/teams/COL/2018-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='2',url))
end<-as.numeric(start)+3
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,start,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2018<-df
```

2017

```
url<-"https://www.baseball-reference.com/teams/COL/2017-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='2',url))
end<-as.numeric(start)+3
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
```

```
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,start,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2017<-df
#View(df2017)
```

‘

2016

```
url<-"https://www.baseball-reference.com/teams/COL/2016-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='2',url))
end<-as.numeric(start)+3
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,start,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2016<-df
#View(df2016)
```

2015

```
url<-"https://www.baseball-reference.com/teams/COL/2015-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='2',url))
end<-as.numeric(start)+3
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,start,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
```

```
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2015<-df
#View(df2015)
```

2014

```
url<-"https://www.baseball-reference.com/teams/COL/2014-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='2',url))
end<-as.numeric(start)+3
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,start,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
```

```
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2014<-df
#View(df2014)
```

2013

```
url<-"https://www.baseball-reference.com/teams/COL/2013-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='2',url))
end<-as.numeric(start)+3
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,start,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
```

```
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2013<-df
#View(df2013)
```

2012

```
url<-"https://www.baseball-reference.com/teams/COL/2012-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
```

```
df<-select(df,-11,-12)
df2012<-df
```

2011

```
url<-"https://www.baseball-reference.com/teams/COL/2011-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='2',url))

end<-as.numeric(start)+3
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,start,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2011<-df
#View(df2011)
```

2010

```
url<-"https://www.baseball-reference.com/teams/COL/2010-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2010<-df
#View(df2010)
```

2009

```
url<-"https://www.baseball-reference.com/teams/COL/2009-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2009<-df
#View(df2009)
```

2008

```
url<-"https://www.baseball-reference.com/teams/COL/2008-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
```

```
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2008<-df
#View(df2008)
```

2007

```
url<-"https://www.baseball-reference.com/teams/COL/2007-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2007<-df
#View(df2007)
```

2006

```
url<-"https://www.baseball-reference.com/teams/COL/2006-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
```

```
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2006<-df
#View(df2006)
```

2005

```
url<-"https://www.baseball-reference.com/teams/COL/2005-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
```

```
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2005<-df
#View(df2005)
```

2004

```
url<-"https://www.baseball-reference.com/teams/COL/2004-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
```

```
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2004<-df
#View(df2004)
```

2003

```
url<-"https://www.baseball-reference.com/teams/COL/2003-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
```

```
df<-select(df,-11,-12)
df2003<-df
#View(df2003)
```

2002

```
url<-"https://www.baseball-reference.com/teams/COL/2002-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2002<-df
#View(df2002)
```

2001

```
url<-"https://www.baseball-reference.com/teams/COL/2001-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2001<-df
#View(df2001)
```

2000

```
url<-"https://www.baseball-reference.com/teams/COL/2000-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/2',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df2000<-df
#View(df2000)
```

1999

```
url<-"https://www.baseball-reference.com/teams/COL/1999-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/1',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
```

```
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df1999<-df
#View(df1999)
```

1998

```
url<-"https://www.baseball-reference.com/teams/COL/1998-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/1',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df1998<-df
#View(df1998)
```

1997

```
url<-"https://www.baseball-reference.com/teams/COL/1997-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/1',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
```

```
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df1997<-df
#View(df1997)
```

1996

```
url<-"https://www.baseball-reference.com/teams/COL/1996-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/1',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
```

```
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year"))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df1996<-df
#View(df1996)
```

1995

```
url<-"https://www.baseball-reference.com/teams/COL/1995-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/1',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
```

```
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
df<-select(df,-11,-12)
df1995<-df
#View(df1995)
```

1994

```
url<-"https://www.baseball-reference.com/teams/COL/1994-schedule-scores.shtml#team_schedule::none"
batting.CO <- read_html(url)
#gets my data into a table form
tables.CO <- batting.CO %>% html_nodes('table') %>% html_table()
##View(tables.CO)
```

```
#string function get my year to be added as an individual column later
start<-(gregexpr(pattern ='/1',url))
end<-as.numeric(start)+4
#end<-(gregexpr(pattern = (start+4),url))

##View(start)
##View(end)
#pulls my year so that i can add it, data set does not denote year in columns
year<-substr(url,as.numeric(start)+1,end)
#year
```

```
#currently my table is in an html table this step turns it into a usable table
df <- data.frame(matrix(unlist(tables.CO), nrow=21, byrow=T),stringsAsFactors=FALSE)
#in the data there are columns which contain only meta data, I  remove these columns in this step
df<-df[,!(df[1,]=='Gm#')]
##View(df)
#Row 3 is in consequential so I remove it from my data
df<-slice(df,-3)
#once i remove row 3 I want only the first 8 rows and row 17
df<-slice(df,0:8,17)
#my data is wide wth a column for each game, transpose will flip this and make it a row per game
df<-transpose(df)
#add my column headings
names(df)[0:9]=c("Gm#","Date","Tm","Home_Away","Opp","W/L","R","RA","Day_Night")
#I need to calculate total runs R and RA came in as chars need to change to create the equation
df$R=as.numeric(df$R)
df$RA=as.numeric(df$RA)
df<- df%>% mutate(Total=R+RA)
df<- df%>% mutate(Year=year)
df<- df%>% mutate(Space=", ")

#Year was not listed in the date column so I will bring it
df$Date<- with(df, paste0(Date,Space,Year))
# df<-unite_("Date",c("Date","","space",Year))
#remove the excess columns created in adding the year to my date (",", "YEar")
```

```r
df<-select(df,-11,-12)
df1994<-df
#View(df1994)
```

```r
baseball_data<-rbind(df2018,    df2017, df2016, df2015, df2014, df2013, df2012, df2011, df2010, df2009,
##View(baseball_data)

baseball_data<-baseball_data%>%mutate(Day_Type=(substr(baseball_data$Date,0,(as.numeric(gregexpr(patter


baseball_data<-baseball_data%>%mutate(Date=(substr(baseball_data$Date,(as.numeric(gregexpr(pattern ='y,
head(baseball_data)
```

```
##   Gm#          Date  Tm Home_Away Opp W/L R RA Day_Night Total  Day_Type
## 1   1  Mar 29, 2018 COL         @ ARI   L 2  8         N    10  Thursday
## 2   2  Mar 30, 2018 COL         @ ARI   L 8  9         N    17    Friday
## 3   3  Mar 31, 2018 COL         @ ARI   W 2  1         N     3  Saturday
## 4   4   Apr 2, 2018 COL         @ SDP   W 7  4         N    11    Monday
## 5   5   Apr 3, 2018 COL         @ SDP   L 4  8         N    12   Tuesday
## 6   6   Apr 4, 2018 COL         @ SDP   W 5  2         N     7 Wednesday
```

```r
#I will write the data into CSV to work with it in tableau
#write.csv(baseball_data, "C:/Users/EricBreton/Desktop/applied ML/baseball_Data.csv")
#View(babse)
```