

# DataCleaning.pdf

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

```
# installing packages needed
#install.packages("readr")
#install.packages("dplyr")

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
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

```
# importing the CSV
GameStats = read_csv("/Users/hannahbrown/Downloads/Practice1/GameStats.csv")

## New names:
## Rows: 7360 Columns: 25
## -- Column specification
## ----- Delimiter: "," chr
## (5): School, Date, ...4, Opponent, ...6 dbl (20): G#, PassCmp, PassAtt,
## PassPct, PassYds, PassTD, RushAtt, RushYds, ...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * ` ` -> `...4`
## * ` ` -> `...6`

#getting a first look at the data set, looking at variable names and number of rows
head(GameStats)

## # A tibble: 6 x 25
##   School Date   `G#` ...4 Opponent ...6 PassCmp PassAtt PassPct PassYds PassTD
```

```
##   <chr>  <chr> <dbl> <chr> <chr>      <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Georg~ 9/4/~      1 <NA> Charlot~ L          25        43        58.1        299        2
## 2 Michi~ 9/4/~      1 @      Western~ W          15        31        48.4        256        2
## 3 Orego~ 9/4/~      1 <NA> Weber S~ W          12        22        54.5        110        2
## 4 SMU     9/4/~      1 <NA> Baylor  L          16        24        66.7        166        2
## 5 Syrac~ 9/4/~      1 <NA> Rhode I~ W          10        20        50          114        2
## 6 Washi~ 9/4/~      1 @      Boise S~ L          20        35        57.1        150        0
## # i 14 more variables: RushAtt <dbl>, RushYds <dbl>, RushAvg <dbl>,
## #   RushTD <dbl>, XPM <dbl>, XPA <dbl>, XPPercent <dbl>, FGM <dbl>, FGA <dbl>,
## #   FGPercent <dbl>, KickPts <dbl>, Fum <dbl>, Int <dbl>, TotalT0 <dbl>
```

```
names(GameStats)
```

```
## [1] "School"      "Date"        "G#"          "...4"        "Opponent"    "...6"
## [7] "PassCmp"     "PassAtt"     "PassPct"     "PassYds"     "PassTD"      "RushAtt"
## [13] "RushYds"     "RushAvg"     "RushTD"      "XPM"         "XPA"         "XPPercent"
## [19] "FGM"         "FGA"         "FGPercent"   "KickPts"     "Fum"         "Int"
## [25] "TotalT0"
```

```
nrow(GameStats)
```

```
## [1] 7360
```

```
# looking at the unique values of the variable School
```

```
x <- unique(GameStats$School)
```

```
# looking at the unique values of the Opponent variable to see what schools need to be renamed for easi
```

```
y <- unique(GameStats$Opponent)
```

```
#arranging the games by date to make it easier to see the corresponding rows for each game
```

```
GameStats <- GameStats %>% arrange(Date)
```

```
head(GameStats)
```

```
## # A tibble: 6 x 25
##   School Date   `G#` ...4 Opponent ...6 PassCmp PassAtt PassPct PassYds PassTD
##   <chr>  <chr> <dbl> <chr> <chr>      <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Flori~ 1/1/~      14 N      Michigan L          10        25        40          155        1
## 2 Iowa   1/1/~      14 N      Stanford L          21        33        63.6        239        2
## 3 Michi~ 1/1/~      13 N      Florida  W          20        31        64.5        278        3
## 4 North~ 1/1/~      13 N      Tenness~ L          14        33        42.4        129        0
## 5 Notre~ 1/1/~      13 N      Ohio St~ L          22        38        57.9        284        2
## 6 Ohio ~ 1/1/~      13 N      Notre D~ W          19        31        61.3        211        1
## # i 14 more variables: RushAtt <dbl>, RushYds <dbl>, RushAvg <dbl>,
## #   RushTD <dbl>, XPM <dbl>, XPA <dbl>, XPPercent <dbl>, FGM <dbl>, FGA <dbl>,
## #   FGPercent <dbl>, KickPts <dbl>, Fum <dbl>, Int <dbl>, TotalT0 <dbl>
```

```
# One Row For Each Game
```

```
# renaming Schools in the Opponent column to match the names in the School column
```

```
GameStats <- GameStats %>% mutate(Opponent = ifelse(Opponent == "Texas-San Antonio", "UTSA", Opponent))
```

```
#creating a data frame with all of the away games, looking at the number of rows
```

```
awaygames <- GameStats %>% filter(...4 == "@")
```

```
nrow(awaygames)
```

```
## [1] 3156
```

```
head(awaygames,50)
```

```
## # A tibble: 50 x 25
##   School      Date `G#` ...4 Opponent ...6 PassCmp PassAtt PassPct PassYds
##   <chr>      <chr> <dbl> <chr> <chr>      <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Miami (FL) 10/1~    4 @    Cincinn~ L          25         40      62.5      281
## 2 Oklahoma   10/1~    4 @    Texas C~ W          23         30      76.7      274
## 3 Old Dominion 10/1~    5 @    Charlot~ W          17         29      58.6      274
## 4 Oregon     10/1~    5 @    Washing~ L          17         27       63       202
## 5 Oregon State 10/1~    4 @    Colorado L          13         32      40.6      100
## 6 Purdue     10/1~    4 @    Maryland L          22         53      41.5      195
## 7 Rice       10/1~    5 @    Souther~ L          16         20       80       166
## 8 Rutgers    10/1~    5 @    Ohio St~ L           3         16      18.8       33
## 9 San Diego S~ 10/1~    4 @    South A~ L          13         22      59.1      159
## 10 San Jose St~ 10/1~    5 @    New Mex~ L          16         30      53.3      314
## # i 40 more rows
## # i 15 more variables: PassTD <dbl>, RushAtt <dbl>, RushYds <dbl>,
## #   RushAvg <dbl>, RushTD <dbl>, XPM <dbl>, XPA <dbl>, XPPercent <dbl>,
## #   FGM <dbl>, FGA <dbl>, FGPercent <dbl>, KickPts <dbl>, Fum <dbl>, Int <dbl>,
## #   TotalTO <dbl>
```

```
# again looking at the unique values in the School column
```

```
unique(GameStats$School)
```

```
##   [1] "Florida"           "Iowa"              "Michigan"
##   [4] "Northwestern"     "Notre Dame"        "Ohio State"
##   [7] "Oklahoma State"   "Ole Miss"          "Stanford"
##  [10] "Tennessee"        "Alabama"           "Auburn"
##  [13] "Clemson"          "Georgia"           "LSU"
##  [16] "Oklahoma"         "South Carolina"    "UCF"
##  [19] "Kentucky"         "Mississippi State" "Penn State"
##  [22] "Texas"            "Washington"        "Arizona State"
##  [25] "Arkansas"         "Kansas State"      "Oregon"
##  [28] "Texas Christian"  "West Virginia"     "USC"
##  [31] "Western Michigan" "Wisconsin"         "Cincinnati"
##  [34] "Miami (FL)"       "Old Dominion"      "Oregon State"
##  [37] "Pitt"             "Purdue"            "Rice"
##  [40] "Rutgers"          "San Diego State"   "San Jose State"
##  [43] "SMU"              "South Alabama"     "South Florida"
##  [46] "Southern Mississippi" "Syracuse"         "Temple"
##  [49] "Texas A&M"        "Texas State"       "Troy"
##  [52] "Tulane"           "UCLA"              "UNLV"
##  [55] "Utah"             "Utah State"        "UTEP"
##  [58] "Vanderbilt"       "Virginia"           "Wake Forest"
##  [61] "Washington State" "Western Kentucky"  "Wyoming"
##  [64] "Air Force"        "Akron"              "Appalachian State"
##  [67] "Arizona"          "Ball State"         "Baylor"
##  [70] "Boise State"      "Boston College"    "Bowling Green State"
##  [73] "Buffalo"          "California"         "Central Michigan"
##  [76] "Charlotte"        "Colorado"           "Colorado State"
##  [79] "Duke"             "East Carolina"     "Eastern Michigan"
##  [82] "Florida Atlantic" "Florida International" "Florida State"
##  [85] "Fresno State"     "Georgia State"      "Georgia Tech"
##  [88] "Hawaii"           "Idaho"              "Illinois"
```

## [91]	"Indiana"	"Iowa State"	"Kent State"
## [94]	"Louisiana"	"Louisiana Tech"	"Louisiana-Monroe"
## [97]	"Louisville"	"Marshall"	"Maryland"
## [100]	"Massachusetts"	"Memphis"	"Miami (OH)"
## [103]	"Michigan State"	"Middle Tennessee State"	"Minnesota"
## [106]	"Missouri"	"Navy"	"Nebraska"
## [109]	"Nevada"	"New Mexico"	"New Mexico State"
## [112]	"North Carolina"	"North Carolina State"	"North Texas"
## [115]	"Northern Illinois"	"Ohio"	"Texas Tech"
## [118]	"Toledo"	"Tulsa"	"UTSA"
## [121]	"Army"	"Brigham Young"	"Connecticut"
## [124]	"Kansas"	"Georgia Southern"	"Coastal Carolina"
## [127]	"Houston"	"UAB"	"Virginia Tech"
## [130]	"Arkansas State"	"Liberty"	

```
# again looking at the unique values in the Opponent column
unique(GameStats$Opponent)
```

## [1]	"Michigan"	"Stanford"
## [3]	"Florida"	"Tennessee"
## [5]	"Ohio State"	"Notre Dame"
## [7]	"Ole Miss"	"Oklahoma State"
## [9]	"Iowa"	"Northwestern"
## [11]	"Clemson"	"UCF"
## [13]	"Alabama"	"Oklahoma"
## [15]	"South Carolina"	"LSU"
## [17]	"Georgia"	"Auburn"
## [19]	"Texas"	"Mississippi State"
## [21]	"Penn State"	"Washington"
## [23]	"Kentucky"	"West Virginia"
## [25]	"Kansas State"	"Arkansas"
## [27]	"Texas Christian"	"Oregon"
## [29]	"Arizona State"	"USC"
## [31]	"Wisconsin"	"Western Michigan"
## [33]	"Miami (FL)"	"Cincinnati"
## [35]	"Charlotte"	"Memphis"
## [37]	"Washington State"	"Colorado"
## [39]	"Minnesota"	"Marshall"
## [41]	"Maryland"	"Southern Mississippi"
## [43]	"South Alabama"	"New Mexico"
## [45]	"Temple"	"San Diego State"
## [47]	"Texas A&M"	"Rice"
## [49]	"SMU"	"Incarnate Word"
## [51]	"Idaho"	"Massachusetts"
## [53]	"East Carolina"	"Arizona"
## [55]	"Fresno State"	"California"
## [57]	"Boise State"	"Louisiana Tech"
## [59]	"Duke"	"North Carolina State"
## [61]	"Houston Baptist"	"Central Michigan"
## [63]	"Colorado State"	"Navy"
## [65]	"Kent State"	"Georgia State"
## [67]	"UCLA"	"Alcorn State"
## [69]	"Louisiana-Monroe"	"Northern Illinois"
## [71]	"Iowa State"	"Utah State"
## [73]	"Buffalo"	"Eastern Michigan"

## [75]	"Boston College"	"Utah"
## [77]	"Old Dominion"	"South Florida"
## [79]	"Louisville"	"Oregon State"
## [81]	"Wyoming"	"Virginia"
## [83]	"Bowling Green State"	"Vanderbilt"
## [85]	"Florida International"	"Florida Atlantic"
## [87]	"North Carolina"	"UNLV"
## [89]	"Appalachian State"	"Nevada"
## [91]	"Troy"	"Nebraska"
## [93]	"Michigan State"	"Baylor"
## [95]	"Akron"	"New Mexico State"
## [97]	"UTEP"	"Missouri"
## [99]	"Pitt"	"Purdue"
## [101]	"Tulane"	"Georgia Tech"
## [103]	"Ohio"	"Indiana"
## [105]	"North Texas"	"Air Force"
## [107]	"Illinois"	"Hawaii"
## [109]	"San Jose State"	"Louisiana"
## [111]	"Florida State"	"Wake Forest"
## [113]	"Middle Tennessee State"	"Ball State"
## [115]	"Syracuse"	"Miami (OH)"
## [117]	"Rutgers"	"Connecticut"
## [119]	"Kansas"	"Army"
## [121]	"Brigham Young"	"Texas Tech"
## [123]	"Toledo"	"Texas State"
## [125]	"UTSA"	"Tulsa"
## [127]	"Western Kentucky"	"Portland State"
## [129]	"Georgia Southern"	"Houston"
## [131]	"Coastal Carolina"	"UAB"
## [133]	"Rhode Island"	"Arkansas State"
## [135]	"Alabama State"	"Liberty"
## [137]	"Virginia Tech"	"Lafayette"
## [139]	"Bucknell"	"Idaho State"
## [141]	"Tennessee-Martin"	"Prairie View A&M"
## [143]	"Tennessee State"	"Wagner"
## [145]	"Bethune-Cookman"	"Austin Peay"
## [147]	"Samford"	"Eastern Kentucky"
## [149]	"Charleston Southern"	"Maine"
## [151]	"Southern Utah"	"Western Carolina"
## [153]	"Virginia Military Institute"	"Chattanooga"
## [155]	"Citadel"	"Colgate"
## [157]	"Wofford"	"Mercer"
## [159]	"Delaware State"	"Presbyterian"
## [161]	"Morgan State"	"Alabama A&M"
## [163]	"Tennessee Tech"	"Drake"
## [165]	"Norfolk State"	"Duquesne"
## [167]	"Robert Morris"	"Albany"
## [169]	"Gardner-Webb"	"South Dakota State"
## [171]	"Northern Colorado"	"Florida A&M"
## [173]	"Central Arkansas"	"Central Connecticut State"
## [175]	"Kennesaw State"	"Southeastern Louisiana"
## [177]	"Missouri State"	"California-Davis"
## [179]	"Northwestern State"	"Savannah State"
## [181]	"Weber State"	"Holy Cross"

```
## [183] "Sacramento State"      "Elon"
## [185] "North Dakota"          "Monmouth"
## [187] "Sam Houston State"      "Abilene Christian"
## [189] "Illinois State"         "Arkansas-Pine Bluff"
## [191] "Montana State"          "Campbell"
## [193] "Eastern Washington"     "James Madison"
## [195] "East Tennessee State"   "Stephen F. Austin"
## [197] "Northern Iowa"          "Indiana State"
## [199] "Nicholls State"         "Grambling State"
## [201] "Howard"                 "South Dakota"
## [203] "William & Mary"          "Jackson State"
## [205] "Fordham"                "Villanova"
## [207] "Southern"               "Northern Arizona"
## [209] "Richmond"              "Stony Brook"
## [211] "Eastern Illinois"       "Southeast Missouri State"
## [213] "Furman"                 "South Carolina State"
## [215] "Lamar"                  "North Carolina A&T"
## [217] "McNeese State"          "Jacksonville State"
## [219] "Youngstown State"       "North Carolina Central"
## [221] "Western Illinois"       "Murray State"
## [223] "Cal Poly"               "New Hampshire"
## [225] "Texas Southern"         "Montana"
## [227] "Lehigh"                 "Delaware"
## [229] "North Dakota State"     "Mississippi Valley State"
## [231] "Hampton"                "Southern Illinois"
## [233] "Towson"                 "San Diego"
```

```
# creating a data frame with the stats of all of the home teams, checking the number of rows for home t
home <- GameStats %>% filter(is.na(...4))
nrow(home)
```

```
## [1] 3680
```

```
# creating a data frame with all of the stats for the home and neutral game teams, checking the number
home_neutral_games <- GameStats %>% filter(...4!="@"|is.na(...4))
nrow(home_neutral_games)
```

```
## [1] 4204
```

```
head(home_neutral_games)
```

```
## # A tibble: 6 x 25
##   School Date `G#` ...4 Opponent ...6 PassCmp PassAtt PassPct PassYds PassTD
##   <chr> <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Flori~ 1/1/~ 14 N Michigan L 10 25 40 155 1
## 2 Iowa 1/1/~ 14 N Stanford L 21 33 63.6 239 2
## 3 Michi~ 1/1/~ 13 N Florida W 20 31 64.5 278 3
## 4 North~ 1/1/~ 13 N Tenness~ L 14 33 42.4 129 0
## 5 Notre~ 1/1/~ 13 N Ohio St~ L 22 38 57.9 284 2
## 6 Ohio ~ 1/1/~ 13 N Notre D~ W 19 31 61.3 211 1
## # i 14 more variables: RushAtt <dbl>, RushYds <dbl>, RushAvg <dbl>,
## # RushTD <dbl>, XPM <dbl>, XPA <dbl>, XPPercent <dbl>, FGM <dbl>, FGA <dbl>,
## # FGPercent <dbl>, KickPts <dbl>, Fum <dbl>, Int <dbl>, TotalTO <dbl>
```

```
# creating a data frame with all of the neutral games, checking the number of neutral games
neutral <- GameStats %>% filter(...4=="N")
```

```

nrow(neutral)

## [1] 524

head(neutral)

## # A tibble: 6 x 25
##   School Date `G#` ...4 Opponent ...6 PassCmp PassAtt PassPct PassYds PassTD
##   <chr> <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Flori~ 1/1/~ 14 N Michigan L 10 25 40 155 1
## 2 Iowa 1/1/~ 14 N Stanford L 21 33 63.6 239 2
## 3 Michi~ 1/1/~ 13 N Florida W 20 31 64.5 278 3
## 4 North~ 1/1/~ 13 N Tenness~ L 14 33 42.4 129 0
## 5 Notre~ 1/1/~ 13 N Ohio St~ L 22 38 57.9 284 2
## 6 Ohio ~ 1/1/~ 13 N Notre D~ W 19 31 61.3 211 1
## # i 14 more variables: RushAtt <dbl>, RushYds <dbl>, RushAvg <dbl>,
## # RushTD <dbl>, XPM <dbl>, XPA <dbl>, XPPercent <dbl>, FGM <dbl>, FGA <dbl>,
## # FGPercent <dbl>, KickPts <dbl>, Fum <dbl>, Int <dbl>, TotalTO <dbl>

# creating a data frame with all of the stats for the neutral and away game teams
neutral_away_games <- GameStats %>% filter(!is.na(...4))

nrow(neutral_away_games)

## [1] 3680

head(neutral_away_games, 20)

## # A tibble: 20 x 25
##   School Date `G#` ...4 Opponent ...6 PassCmp PassAtt PassPct PassYds
##   <chr> <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>
## 1 Florida 1/1/~ 14 N Michigan L 10 25 40 155
## 2 Iowa 1/1/~ 14 N Stanford L 21 33 63.6 239
## 3 Michigan 1/1/~ 13 N Florida W 20 31 64.5 278
## 4 Northwestern 1/1/~ 13 N Tenness~ L 14 33 42.4 129
## 5 Notre Dame 1/1/~ 13 N Ohio St~ L 22 38 57.9 284
## 6 Ohio State 1/1/~ 13 N Notre D~ W 19 31 61.3 211
## 7 Oklahoma St~ 1/1/~ 13 N Ole Miss L 27 45 60 303
## 8 Ole Miss 1/1/~ 13 N Oklahom~ W 22 35 62.9 347
## 9 Stanford 1/1/~ 14 N Iowa W 12 21 57.1 223
## 10 Tennessee 1/1/~ 13 N Northwe~ W 16 30 53.3 194
## 11 Alabama 1/1/~ 13 N Clemson W 16 24 66.7 120
## 12 Auburn 1/1/~ 14 N UCF L 28 43 65.1 331
## 13 Clemson 1/1/~ 14 N Alabama L 18 37 48.6 124
## 14 Georgia 1/1/~ 14 N Oklahoma W 20 29 69 210
## 15 LSU 1/1/~ 13 N Notre D~ L 19 33 57.6 229
## 16 Michigan 1/1/~ 13 N South C~ L 21 45 46.7 203
## 17 Notre Dame 1/1/~ 13 N LSU W 17 28 60.7 216
## 18 Oklahoma 1/1/~ 14 N Georgia L 24 36 66.7 289
## 19 South Carol~ 1/1/~ 13 N Michigan W 19 32 59.4 239
## 20 UCF 1/1/~ 13 N Auburn W 16 35 45.7 242
## # i 15 more variables: PassTD <dbl>, RushAtt <dbl>, RushYds <dbl>,
## # RushAvg <dbl>, RushTD <dbl>, XPM <dbl>, XPA <dbl>, XPPercent <dbl>,
## # FGM <dbl>, FGA <dbl>, FGPercent <dbl>, KickPts <dbl>, Fum <dbl>, Int <dbl>,
## # TotalTO <dbl>

```

```

# merging the home, away, and neutral game datasets together so there is one row for each game
joined <- home_neutral_games %>% full_join(neutral_away_games, c("School" = "Opponent", "Date"))

nrow(joined)

## [1] 4209

# Remove games played on a neutral field (value "N" in the 4th column)
noneutral <- joined %>%
  filter(...4.y!="N")

nrow(noneutral)

## [1] 3156

head(noneutral,50)

## # A tibble: 50 x 48
##   School      Date `G#.x` ...4.x Opponent ...6.x PassCmp.x PassAtt.x PassPct.x
##   <chr>      <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl>
## 1 Cincinnati 10/1~ 5 <NA> Miami (~ W 22 33 66.7
## 2 Oklahoma S~ 10/1~ 5 <NA> Texas W 19 28 67.9
## 3 Ole Miss 10/1~ 5 <NA> Memphis W 30 44 68.2
## 4 Penn State 10/1~ 5 <NA> Minneso~ W 19 42 45.2
## 5 Pitt 10/1~ 5 <NA> Marshall W 17 24 70.8
## 6 South Alab~ 10/1~ 5 <NA> San Die~ W 16 21 76.2
## 7 South Caro~ 10/1~ 5 <NA> Texas A~ L 22 35 62.9
## 8 Southern M~ 10/1~ 5 <NA> Rice W 25 36 69.4
## 9 Temple 10/1~ 5 <NA> SMU W 7 18 38.9
## 10 Texas Chri~ 10/1~ 5 <NA> Oklahoma L 26 45 57.8
## # i 40 more rows
## # i 39 more variables: PassYds.x <dbl>, PassTD.x <dbl>, RushAtt.x <dbl>,
## # RushYds.x <dbl>, RushAvg.x <dbl>, RushTD.x <dbl>, XPM.x <dbl>, XPA.x <dbl>,
## # XPPercent.x <dbl>, FGM.x <dbl>, FGA.x <dbl>, FGPercent.x <dbl>,
## # KickPts.x <dbl>, Fum.x <dbl>, Int.x <dbl>, TotalTO.x <dbl>, School.y <chr>,
## # `G#.y` <dbl>, ...4.y <chr>, ...6.y <chr>, PassCmp.y <dbl>, PassAtt.y <dbl>,
## # PassPct.y <dbl>, PassYds.y <dbl>, PassTD.y <dbl>, RushAtt.y <dbl>, ...

## Data Cleaning: Part C
# First 3 variables in the dataset are Date, Home, and Away
clean_data_c <- noneutral %>% relocate(Date, .before = School) %>% relocate(Opponent, .after = School) %>%
  rename(Home = School) %>%
  rename(Away = Opponent)

## Data Cleaning: Part D
# Creating the HomeWins binary Variable with a 1 if the home team won, and a 0 if otherwise

clean_data_d <- clean_data_c %>% mutate(HomeWins = ifelse(...6.x=="W", 1, 0)) %>%
  relocate(HomeWins, .after = Away)

## Data Cleaning: Part E
#Creating 2 Variables for Each Team Statistic
# Renaming variables to indicate if they are stats for the Home or Away Team, Putting them in order

clean_data_e <- clean_data_d %>%
  rename(HPassCmp = PassCmp.x) %>% relocate(HPassCmp, .after = HomeWins) %>%

```



```

rename(APassCmp = PassCmp.y) %>% relocate(APassCmp, .after = HPassCmp) %>%
rename(HPassAtt = PassAtt.x) %>% relocate(HPassAtt, .after = APassCmp) %>%
rename(APassAtt = PassAtt.y) %>% relocate(APassAtt, .after = HPassAtt) %>%
rename(HPassPct = PassPct.x) %>% relocate(HPassPct, .after = APassAtt) %>%
rename(APassPct = PassPct.y) %>% relocate(APassPct, .after = HPassPct) %>%
rename(HPassYds = PassYds.x) %>% relocate(HPassYds, .after = APassPct) %>%
rename(APassYds = PassYds.y) %>% relocate(APassYds, .after = HPassYds) %>%
rename(HPassTD = PassTD.x) %>% relocate(HPassTD, .after = APassYds) %>%
rename(APassTD = PassTD.y) %>% relocate(APassTD, .after = HPassTD) %>%
rename(HRushAtt = RushAtt.x) %>% relocate(HRushAtt, .after = APassTD) %>%
rename(ARushAtt = RushAtt.y) %>% relocate(ARushAtt, .after = HRushAtt) %>%
rename(HRushYds = RushYds.x) %>% relocate(HRushYds, .after = ARushAtt) %>%
rename(ARushYds = RushYds.y) %>% relocate(ARushYds, .after = HRushYds) %>%
rename(HRushAvg = RushAvg.x) %>% relocate(HRushAvg, .after = ARushYds) %>%
rename(ARushAvg = RushAvg.y) %>% relocate(ARushAvg, .after = HRushAvg) %>%
rename(HRushTD = RushTD.x) %>% relocate(HRushTD, .after = ARushAvg) %>%
rename(ARushTD = RushTD.y) %>% relocate(ARushTD, .after = HRushTD) %>%
rename(HXPM = XPM.x) %>% relocate(HXPM, .after = ARushTD) %>%
rename(AXPM = XPM.y) %>% relocate(AXPM, .after = HXPM) %>%
rename(HXPA = XPA.x) %>% relocate(HXPA, .after = AXPM) %>%
rename(AXPA = XPA.y) %>% relocate(AXPA, .after = HXPA) %>%
rename(HXPPercent = XPPercent.x) %>% relocate(HXPPercent, .after = AXPA) %>%
rename(AXPPercent = XPPercent.y) %>% relocate(AXPPercent, .after = HXPPercent) %>%
rename(HFGM = FGM.x) %>% relocate(HFGM, .after = AXPPercent) %>%
rename(AFGM = FGM.y) %>% relocate(AFGM, .after = HFGM) %>%
rename(HFGA = FGA.x) %>% relocate(HFGA, .after = AFGM) %>%
rename(AFGA = FGA.y) %>% relocate(AFGA, .after = HFGA) %>%
rename(HFGPercent = FGPercent.x) %>% relocate(HFGPercent, .after = AFGA) %>%
rename(AFGPercent = FGPercent.y) %>% relocate(AFGPercent, .after = HFGPercent) %>%
rename(HKickPts = KickPts.x) %>% relocate(HKickPts, .after = AFGPercent) %>%
rename(AKickPts = KickPts.y) %>% relocate(AKickPts, .after = HKickPts) %>%
rename(HFum = Fum.x) %>% relocate(HFum, .after = AKickPts) %>%
rename(AFum = Fum.y) %>% relocate(AFum, .after = HFum) %>%
rename(HInt = Int.x) %>% relocate(HInt, .after = AFum) %>%
rename(AInt = Int.y) %>% relocate(AInt, .after = HInt) %>%
rename(HTotalTO = TotalTO.x) %>% relocate(HTotalTO, .after = AInt) %>%
rename(ATotalTO = TotalTO.y) %>% relocate(ATotalTO, .after = HTotalTO)

```

```

# Removing Unnecessary Variables

```

```

final_clean_data <- clean_data_e %>% select(-"G#.x", -...4.x, -...6.x, -School.y, -"G#.y", -...4.y, -...

```

```

# Writing the final cleaned dataset to my computer and saving as CleanedGameStats.csv
write.csv(final_clean_data, "CleanedGameStats.csv")

```