**Part 1: Cleaning Combine Data**

Data found at <http://nflsavant.com/about.php>

Code

library(tidyr)

library(dplyr)

combine <- combine %>% select(year,name,position,heightfeet,heightinches,weight,fortyyd,twentyss,threecone,vertical,broad,bench,college,picktotal)

combine1<- filter(combine, year>=2005)

combine1<-filter(combine1,year<=2010)

combine1<-filter(combine1,picktotal>0)

combine1<-unite(combine1, heightfeet, heightinches, col = "height", sep = "-")

combine1<-mutate(combine1, sum = fortyyd+twentyss+threecone+vertical+broad+bench)

combine2<-filter(combine1,sum>0)

combine3<-select(combine2,-sum)

Important Steps

* Used select function to capture columns needed from full dataset
* Used filter function to select participants from 2005-2010
* Used filter function to select participants that were drafted
* Used mutate function to make a column with all the stats from each drill added together
* Used filter function to select participants that participated in every drill then removed the sum column

**Part 2: Cleaning Approximate Value Data**

Data found at <https://www.kaggle.com/ronaldjgrafjr/nfl-draft-outcomes>

Code

library(tidyr)

library(dplyr)

nfl\_draft1<-nfl\_draft %>% select(V2, V6,V9,V15,V16)

colnames(nfl\_draft1) <- c('year','name', 'first4av', 'carav', 'drav')

nfl\_draft2<-filter(nfl\_draft1, year>=2005)

nfl\_draft3<-filter(nfl\_draft2, year<=2010)

nfl\_draft4<-nfl\_draft3 %>% select(-year)

Important Steps

* Used the select function to capture columns needed from full dataset
* Used colnames function to change column names
* Used filter function to select players drafted from 2005-2010
* Removed year column since the combine data has a year column

**Part 3: Combining Data Sets**

Code

capstoneprojectdata<-inner\_join(combine3,nfl\_draft4, by = "name")

Used the inner join function to combine data sets using the names of the players.