Project Restart

Ethan Allavarpu

8/2/2020

Contents

1 Import Data

```r
endseason <- read.csv("../data/endseason.csv")
names(endseason)[[1]] <- "Club"
endseason$rank <- 1:20

endseason <- endseason[order(endseason$Club), ]
seasonstop <- read.csv("../data/seasonstop.csv")
names(seasonstop)[[1]] <- "Club"
seasonstop$rank <- 1:20
seasonstop <- seasonstop[order(seasonstop$Club), ]
endstats <- as.matrix(select(endseason, !Club))
stopstats <- as.matrix(select(seasonstop, !Club))
restartstats <- endstats - stopstats
projectrestart <- data.frame(endseason$Club[order(endseason$Club)], restartstats)
names(projectrestart)[[1]] <- "Club"
endseason$avgpts <- endseason %>% summarize(avgpts = P / GP) %>% as.matrix()
```

```
Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
dplyr 1.1.0.
i Please use `reframe()` instead.
i When switching from `summarise()` to `reframe()`, remember that `reframe()`
  always returns an ungrouped data frame and adjust accordingly.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.
```

```r
seasonstop$avgpts <- seasonstop %>% summarize(avgpts = P / GP) %>% as.matrix()
```

```
Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
dplyr 1.1.0.
i Please use `reframe()` instead.
i When switching from `summarise()` to `reframe()`, remember that `reframe()`
  always returns an ungrouped data frame and adjust accordingly.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

```
generated.
projectrestart$avgpts <- projectrestart %>% summarize(avgpts = P / GP) %>% as.matrix()

Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
dplyr 1.1.0.
i Please use `reframe()` instead.
i When switching from `summarise()` to `reframe()`, remember that `reframe()`
  always returns an ungrouped data frame and adjust accordingly.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.
```

2 Visuals

```
club_abbrev <- c("LIV", "MCI", "LEI", "CHE", "MUN", "WOL", "SHU", "TOT",
                 "ARS", "BUR", "CRY", "EVE", "NEW", "SOU", "BHA", "WHU",
                 "WAT", "BOU", "AVL", "NOR")
averages <- data.frame(seasonstop$Club[order(seasonstop$Club)],
                       seasonstop$avgpts,
                       projectrestart$avgpts,
                       endseason$avgpts)
names(averages) <- c("Club", "Start", "Restart", "End")
averages$ptdiff <- averages %>% summarize(ptdiff = Restart - Start) %>% as.matrix()
plot(Restart ~ Start, data = averages, col = rainbow(20)[1:20])
abline(a = 0, b = 1)
a<- lm(Restart ~ Start, data = averages)
abline(a)
```
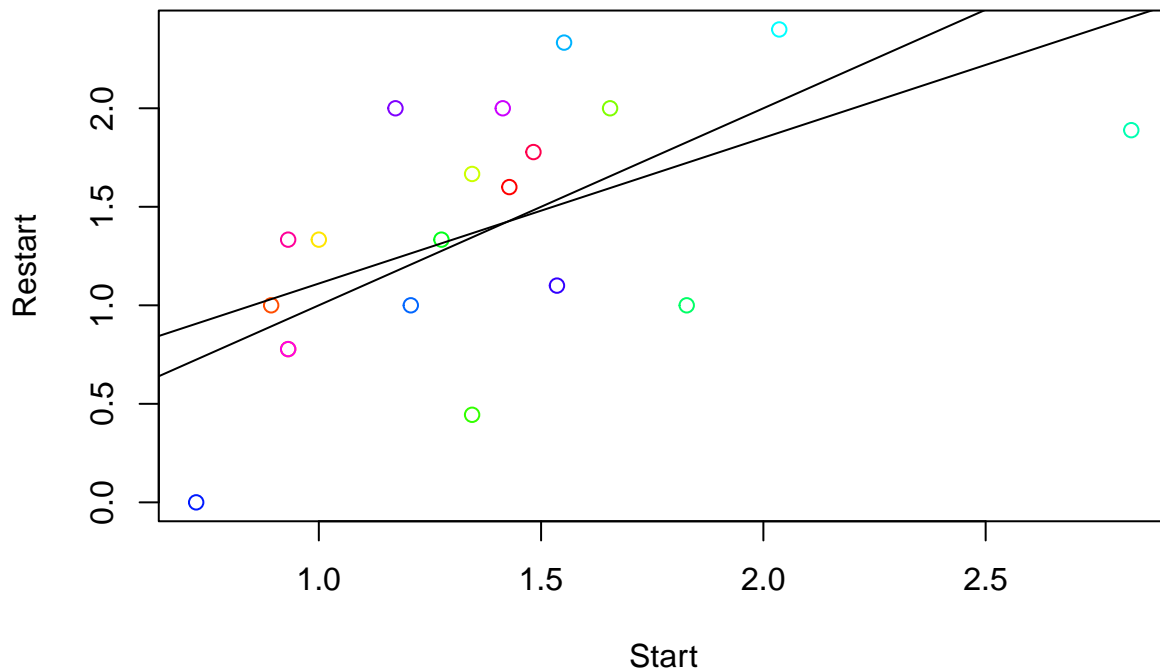


```
plot(ptdiff ~ Start, data = averages, col = rainbow(20)[1:20])
```
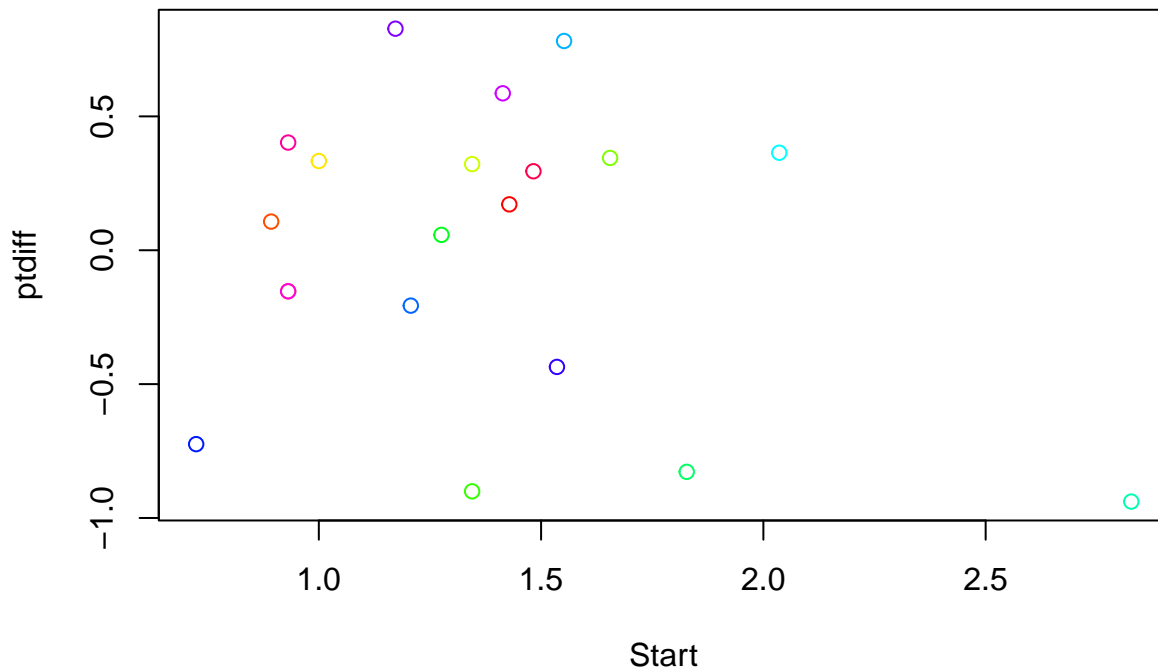
```
summary(aov(abs(ptdiff) ~ Start, data = averages))
```

```
          Df Sum Sq Mean Sq F value Pr(>F)
Start      1 0.2939 0.29390   4.083 0.0585 .
Residuals 18 1.2957 0.07199
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
clubperformance <- chisq.test(projectrestart$P, p = seasonstop$P/sum(seasonstop$P))
clubperformance
```

```
        Chi-squared test for given probabilities

data:  projectrestart$P
X-squared = 36.735, df = 19, p-value = 0.008563
```

```
chisqcontr <- with(clubperformance, (observed - expected)^2 / expected)
names(chisqcontr) <- projectrestart$Club
sort(chisqcontr, decreasing = TRUE)
```

```
             Norwich City           Crystal Palace               Southampton
             6.7878787879             5.8752913753              4.4715166370
           Leicester City                Liverpool           Manchester United
             3.8594970936             3.4086328221              2.8642045455
         Tottenham Hotspur          Manchester City             West Ham United
             1.7006959842             1.6874003190              1.2272727273
 Brighton and Hove Albion                  Arsenal            Sheffield United
             0.7358063393             0.7292929293              0.6046585036
          Newcastle United              Aston Villa                     Burnley
             0.4729527417             0.4558080808              0.4546182984
                  Chelsea              Bournemouth                     Watford
             0.3979640152             0.3418560606              0.3418560606
  Wolverhampton Wanderers                  Everton
```
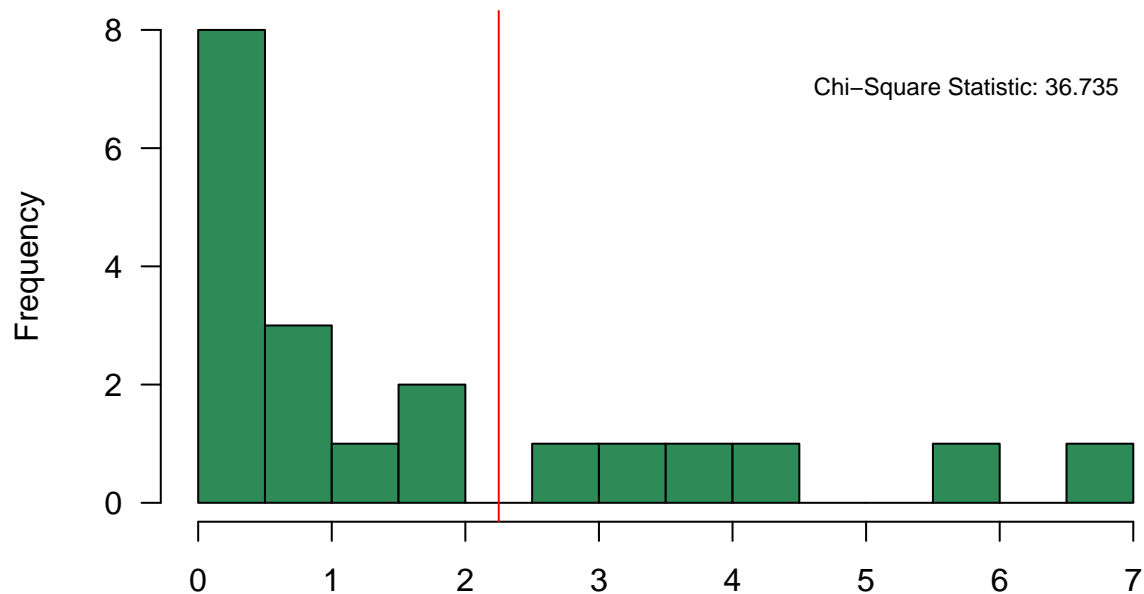
```
            0.3175945502                    0.0001365001
```

```r
hist(chisqcontr,
    breaks = seq(from = 0, to = max(chisqcontr) + 0.5, by = 0.5),
    main = "Chi-Square Contributions by Club",
    xlab = "Size of Chi-Square Contribution",
    col = "seagreen",
    las = 1)
abline(v = 2.25, col = "red")
text(5.75, 7, "Chi-Square Statistic: 36.735", cex = 0.75)
```

Chi−Square Contributions by Club


```r
standings <- seasonstop %>% select(Club, rank)
names(standings)[[2]] <- "Stop"
standings$End <- endseason$rank
standings$Change <- standings %>% summarize(Stop - End) %>% as.matrix() %>% as.numeric()
standings
```

	Club	Stop	End	Change
9	Arsenal	9	8	1
19	Aston Villa	19	17	2
18	Bournemouth	18	18	0
15	Brighton and Hove Albion	15	15	0
10	Burnley	10	10	0
4	Chelsea	4	4	0
11	Crystal Palace	11	14	-3
12	Everton	12	12	0
3	Leicester City	3	5	-2
1	Liverpool	1	1	0
2	Manchester City	2	2	0
5	Manchester United	5	3	2

	Club	Stop	End	Change
13	Newcastle United	13	13	0
20	Norwich City	20	20	0
7	Sheffield United	7	9	-2
14	Southampton	14	11	3
8	Tottenham Hotspur	8	6	2
17	Watford	17	19	-2
16	West Ham United	16	16	0
6	Wolverhampton Wanderers	6	7	-1

```
sum(abs(standings$Change))
```

```
[1] 20
```

```
standings <- standings[order(standings$Stop), ]
movement_col <- c(rgb(0, 0, 0, alpha = 0.5),
                  rgb(0.5, 0, 0, alpha = 0.5),
                  rgb(0, 0.5, 0, alpha = 0.5))
club_abbrev <- c("LIV", "MCI", "LEI", "CHE", "MUN", "WOL", "SHU", "TOT",
                 "ARS", "BUR", "CRY", "EVE", "NEW", "SOU", "BHA", "WHU",
                 "WAT", "BOU", "AVL", "NOR")
plot(1:20, 21 - standings$Stop, yaxt = "n", xaxt = "n", main = "Standings Before and After\nProject Rest
     ylab = "",
     col = movement_col[(standings$Change != 0) + (standings$Change > 0) + 1], pch = 19)
arrows(1:20, 21 - standings$Stop, 1:20, 21 - standings$End, cex = 0.25,
       col = movement_col[(standings$Change != 0) + (standings$Change > 0) + 1], length = 0.1, angle =
points(1:20, 21 - standings$End, col = movement_col[(standings$Change != 0) + (standings$Change > 0) +
abline(h = 1:20, col = rgb(0, 0, 0, alpha = 0.1))
standings$abbreviations <- club_abbrev
axis(1, at = 1:20, labels = club_abbrev, las = 2, cex = 0.5)
abline(h = seq(from = 21, to = 21 - 4.5, by = -0.1), col = rgb(0,1,0, alpha = 0.2))
abline(h = seq(from = (21 - 17.5), to = 0, by = -0.1), col = rgb(1, 0, 0, alpha = 0.2))
abline(h = c(21 - 4.5, 21 - 5.5, 21 - 17.5), col = "grey")
abline(h = seq(from = 21 - 4.5, to = 21 - 5.5, by = -0.1), col = rgb(0, 1, 0, alpha = 0.1))
```

Standings Before and After
Project Restart
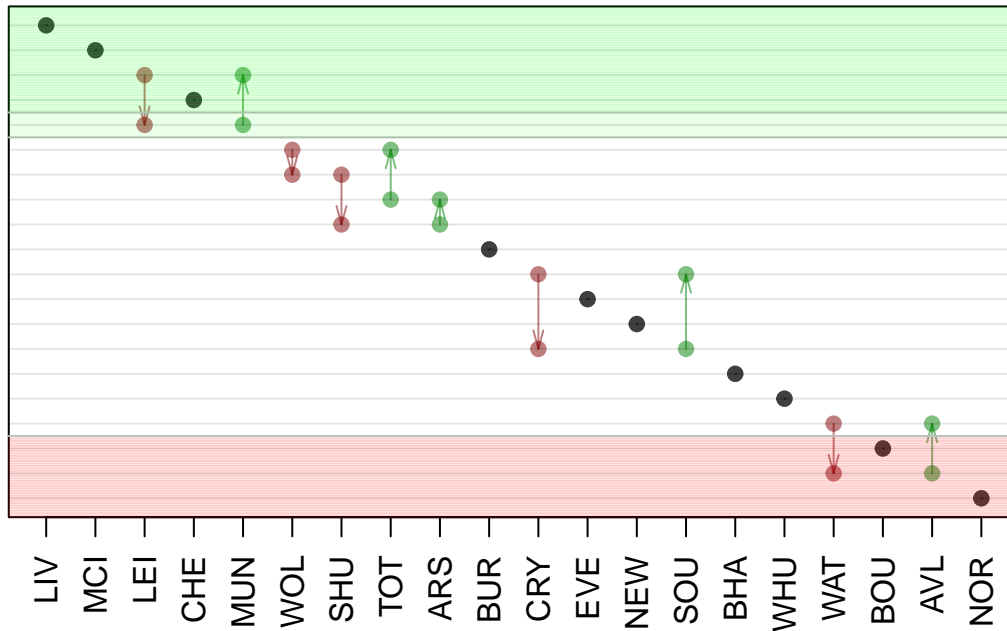

```
club_abbrev <- c("LIV", "MCI", "LEI", "CHE", "MUN", "WOL", "SHU", "TOT",
                 "ARS", "BUR", "CRY", "EVE", "NEW", "SOU", "BHA", "WHU",
                 "WAT", "BOU", "AVL", "NOR")
pointchanges <- seasonstop %>% select(Club, P)
End <- endseason$P
pointchanges <- data.frame(pointchanges, End)
pointchanges <- pointchanges[order(pointchanges$P, decreasing = TRUE), ]
pointchanges[c(6:7, 16, 18), ] <- pointchanges[c(7:6, 18, 16), ]
plot(1:20, pointchanges$P, xaxt = "n", ylab = "Points",
     main = "Total PL Points Before and\nAfter Project Restart", xlab = "", las = 2,
     ylim = c(0, 100), col = rgb(0, 0, 0.5, alpha = 0.25), pch = 19)
points(pointchanges$End, col = rgb(0, 0, 0.5, alpha = 0.5), pch = 19)
arrows(1:20, pointchanges$P, 1:20, pointchanges$End, angle = 15, length = 0.1)
axis(1, 1:20, club_abbrev, las = 2)
```

Total PL Points Before and After Project Restart
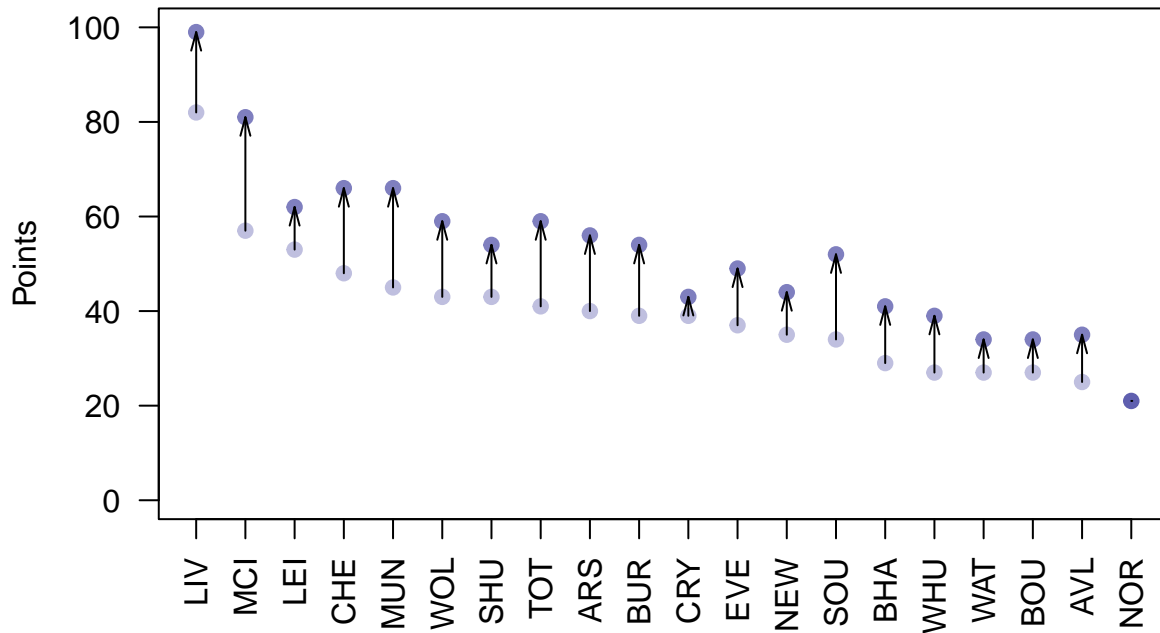

```r
club_abbrev <- c("LIV", "MCI", "LEI", "CHE", "MUN", "WOL", "SHU", "ARS",
                 "TOT", "BUR", "CRY", "EVE", "NEW", "SOU", "BHA", "WHU",
                 "WAT", "BOU", "AVL", "NOR")
averages <- averages[order(averages$Start, decreasing = TRUE), ]
averages[c(6:7, 16, 18), ] <- averages[c(7:6, 18, 16), ]
averages$Abbreviation <- club_abbrev
ppg <- matrix(c(averages$Start, averages$Restart), ncol = 2)
barplot(t(ppg), names.arg = averages$Abbreviation, las = 2,
        ylim = c(0,3),
        main = "Points Earned per Game Throughout\nthe 2019-2020 EPL Season",
        beside = TRUE,
        col = c(rgb(0.5,0,0,0.5),
                rgb(0,0,0.5,0.5)))
legend("topright",
       legend = c("Pre-COVID-19",
                  "Project Restart"),
       fill = c(rgb(0.5,0,0,0.5),
                rgb(0,0,0.5,0.5)),
       inset = 0.05,
       cex = 0.75)
```

Points Earned per Game Throughout the 2019–2020 EPL Season
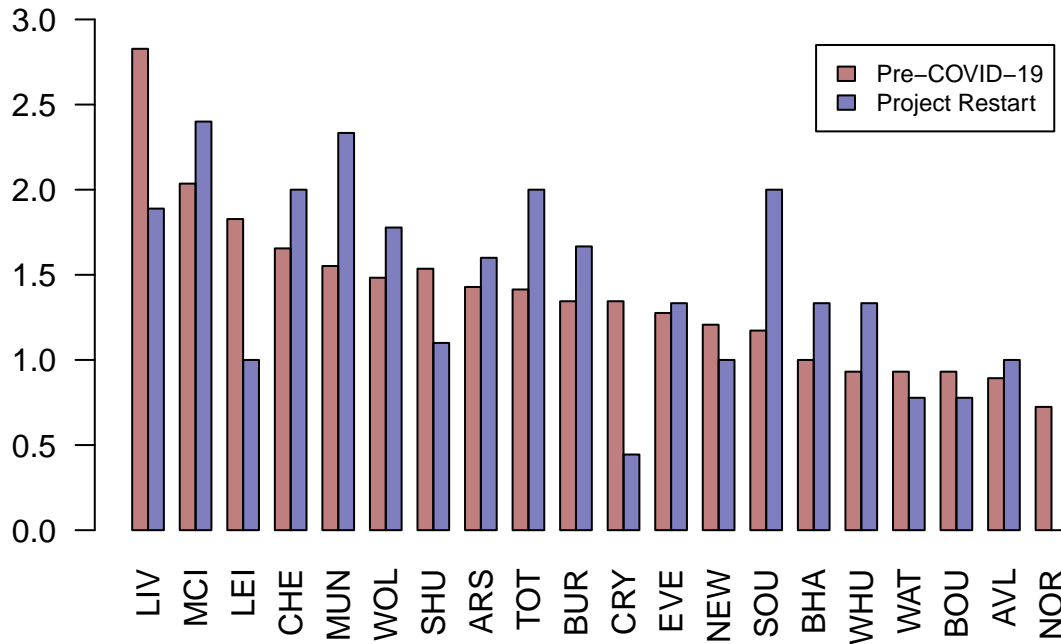
3 Statistical Testing

```r
pre_covid <- data.frame(Club = seasonstop$Club,
                        Successes = seasonstop$P,
                        Trials = seasonstop$GP * 3)
post_covid <- data.frame(Club = projectrestart$Club,
                         Successes = projectrestart$P,
                         Trials = projectrestart$GP * 3)
effects <- data.frame()
for (i in seq_len(nrow(pre_covid))) {
  z_test <- prop.test(c(post_covid[i, 2], pre_covid[i, 2]),
                      c(post_covid[i, 3], pre_covid[i, 3]),
                      conf.level = 1 - 0.05/20)
  effects <-rbind(effects,
                  c(z_test$estimate[1]-z_test$estimate[2],
                  z_test$conf.int[1],
                  z_test$conf.int[2],
                  z_test$p.value))
}
```

```
Warning in prop.test(c(post_covid[i, 2], pre_covid[i, 2]), c(post_covid[i, :
Chi-squared approximation may be incorrect

Warning in prop.test(c(post_covid[i, 2], pre_covid[i, 2]), c(post_covid[i, :
Chi-squared approximation may be incorrect
```

```r
names(effects) <- c("PercentPtsObtainedChange", "lb", "ub", "p_val")
effects <- data.frame(effects)
effects <- data.frame(Club = seasonstop$Club, effects)
```

```
effects <- effects[order(seasonstop$P / seasonstop$GP, decreasing = TRUE), ]
effects[c(6:7, 16, 18), ] <- effects[c(7:6, 18, 16), ]
plot(effects$PercentPtsObtainedChange, ylim = c(-1, 1), xaxt = "n",
     pch = 19, col = rgb(0,0,0, 0.25), xlab = "",
     ylab = "Change in Point Percentage", main = "Shift in Proportion of Points Earned\nAfter Project R
segments(1:20, effects$lb, 1:20, effects$ub, col = "red")
axis(1, at = 1:20, labels = club_abbrev, las = 2)
abline(h = 0, lty = 2, lwd = 1.5)
```


**Shift in Proportion of Points Earned
After Project Restart**
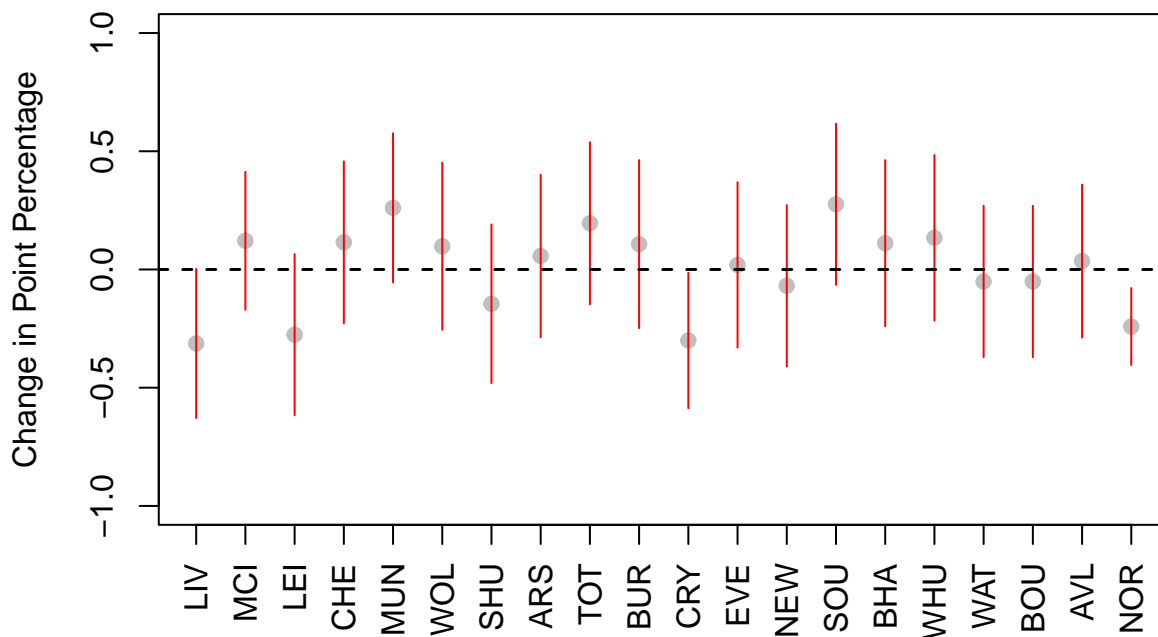
Winners of the Restart: MUN, SOU Losers of the Restart: LIV, LEI, CRY, NOR

```
averages$pctptsbefore <- averages %>% summarize(pctptsbefore = Start / 3) %>% as.matrix() %>% as.numeric
```

```
Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
dplyr 1.1.0.
i Please use `reframe()` instead.
i When switching from `summarise()` to `reframe()`, remember that `reframe()`
  always returns an ungrouped data frame and adjust accordingly.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.
```

```
averages$pctptsafter <- averages %>% summarize(pctptsafter = Restart / 3) %>% as.matrix() %>% as.numeric
```

```
Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
dplyr 1.1.0.
i Please use `reframe()` instead.
i When switching from `summarise()` to `reframe()`, remember that `reframe()`
  always returns an ungrouped data frame and adjust accordingly.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.
```