

Preliminary Thoughts

Some Summarization

```
# group by tier
seasonByTier <- seasonboth %>% group_by(Game.Tier)

# take away rows with promo
seasonByTierWithoutPromotion <- subset(seasonByTier, seasonByTier$Promo == "")

# take the summary
summaryOfNoPromo <- seasonByTierWithoutPromotion %>% summarise(
  tip = mean(Attendance.at.tip),
  total = mean(Total.Attendance)
)

# take only rows with promo
seasonByTierWithPromotion <- subset(seasonByTier, seasonByTier$Promo != "")

# take the summary
summaryOfWithPromo <- seasonByTierWithPromotion %>% summarise(
  tip = mean(Attendance.at.tip),
  total = mean(Total.Attendance)
)

# merge total summary
totalSummary <- merge(summaryOfNoPromo, summaryOfWithPromo, by="Game.Tier")
totalSummary <- rename(totalSummary, "tipWithout" = "tip.x", "totalWithout" = "total.x", "tipWith" = "tip.y", "totalWith" = "total.y")

# create a final summary with percent changes
finalSummary <- totalSummary %>%
  mutate(PercTip = (tipWith/tipWithout - 1)*100) %>%
  mutate(PercTotal = (totalWith/totalWithout - 1)*100)

# see results
finalSummary
```

##	Game.Tier	tipWithout	totalWithout	tipWith	totalWith	PercTip	PercTotal
## 1	Tier A	11520.60	14508.40	14527.50	14972.00	26.100203	3.195390
## 2	Tier B	12060.67	15408.33	15162.75	16466.50	25.720662	6.867496
## 3	Tier C	12335.13	14371.26	13309.43	15435.71	7.898564	7.406820
## 4	Tier D	11517.04	14326.17	10828.56	14026.11	-5.977992	-2.094508

We can see that generally, for higher tier games, there is a much higher amount of people at tip-off (26% and 25% increase for tier A and B respectively). However, there are marginal increases in total attendances for those two tiers (3% and 7% increase), while for tier C, there is an increase of around 7-8% for both tip and

total attendance. For Tier D games, we actually saw a decrease in attendance at tip and total, of around -6 and -2 percent respectively. So perhaps we can argue that promos should not be applied towards games that are already decided to be Tier D.

I would like to not have the data summarized for the next step and be able to check out variances and spread for these different columns.