Duke Offensive Stats: 2022-23

Packages

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.3     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(tidymodels)
```

```
-- Attaching packages --------------------------------------- tidymodels 1.1.1 --
v broom        1.0.5     v rsample      1.2.0
v dials        1.2.0     v tune         1.1.2
v infer        1.0.4     v workflows    1.1.3
v modeldata    1.2.0     v workflowsets 1.0.1
v parsnip      1.1.1     v yardstick    1.2.0
v recipes      1.0.8
-- Conflicts ------------------------------------------ tidymodels_conflicts() --
x scales::discard() masks purrr::discard()
x dplyr::filter()   masks stats::filter()
x recipes::fixed()  masks stringr::fixed()
x dplyr::lag()      masks stats::lag()
```

```
x yardstick::spec() masks readr::spec()
x recipes::step()   masks stats::step()
* Search for functions across packages at https://www.tidymodels.org/find/
```

Home-Game Attendance & Offensive Performance

This section explores if any relationship appears to exist between game attendance and the offensive performance of Duke during games in Wallace Wade Stadium.

Import Data

```
offense_data <- read_csv("data/Duke Stats - DukeOffense.csv")
```

```
Rows: 416 Columns: 40
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (9): OppName, Surface, Day, Site, Result, TV_Coverage, City, State, Type
dbl (25): FPI, FPI_diff, Month, Date, Year, Start_Time, DukePts, OppPts, Poi...
lgl  (6): Rain, 1stSeedQB, SchoolBreak, NatlHoliday, Bowl, UNC_Game

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.
```

```
offense_data <- offense_data |>
  mutate(isHome = if_else(Site == "Home", TRUE, FALSE)) |>
  mutate(Day = as.factor(Day)) |>
  mutate(AttPct = if_else(AttNum/40004 > 1.0, 100.0, AttNum/40004*100))

home_offense_data <- offense_data |>
  filter(isHome == TRUE)

glimpse(home_offense_data)
```

```
Rows: 208
Columns: 41
$ OppName      <chr> "Clemson", "Lafayette", "Northwestern", "Notre Dame", "No~
$ FPI          <dbl> 13.8, NA, 0.8, 20.7, 6.9, -1.7, -0.5, -11.8, NA, -4.0, 6.~
$ FPI_diff     <dbl> 4.8, NA, -8.2, 11.7, -2.1, -10.7, -9.5, -17.1, -5.3, -9.3~
```

```
$ Surface      <chr> "Grass", "Grass", "Grass", "Grass", "Grass", "Grass", "Gr~
$ Month        <dbl> 9, 9, 9, 9, 10, 11, 11, 9, 9, 10, 10, 11, 11, 9, 9, 9, 9,~
$ Date         <dbl> 4, 9, 16, 30, 14, 2, 25, 2, 17, 1, 15, 12, 26, 4, 9, 16, ~
$ Year         <dbl> 2023, 2023, 2023, 2023, 2023, 2023, 2023, 2022, 2022, 202~
$ Day          <fct> Mon, Sat, Sat, Sat, Sat, Thu, Sat, Fri, Sat, Sat, Sat, Sa~
$ Start_Time   <dbl> 20.0, 18.0, 15.5, 19.5, 20.0, 19.5, 12.0, 19.5, 18.0, 19.~
$ Site         <chr> "Home", "Home", "Home", "Home", "Home", "Home", "Home", "~
$ Result       <chr> "W", "W", "W", "L", "W", "W", "W", "W", "W", "W", "L", "W~
$ DukePts      <dbl> 28, 42, 38, 14, 24, 24, 30, 30, 49, 38, 35, 24, 34, 28, 4~
$ OppPts       <dbl> 7, 7, 14, 21, 3, 21, 19, 0, 20, 17, 38, 7, 31, 7, 7, 14, ~
$ PointDiff    <dbl> 21, 35, 24, -7, 21, 3, 11, 30, 29, 21, -3, 17, 3, 21, 35,~
$ AttNum       <dbl> 31638, 17481, 18141, 40768, 31833, 18277, 17639, 20722, 3~
$ AttPct       <dbl> 79.08709, 43.69813, 45.34797, 100.00000, 79.57454, 45.687~
$ ESPN_WinPred <dbl> 0.872, 0.993, 0.698, 0.300, 0.774, 0.812, 0.788, 0.771, N~
$ Rain         <lgl> FALSE, TRUE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, FAL~
$ `1stSeedQB`  <lgl> TRUE, TRUE, TRUE, TRUE, FALSE, FALSE, FALSE, TRUE, TRUE, ~
$ SchoolBreak  <lgl> TRUE, FALSE, FALSE, FALSE, TRUE, FALSE, TRUE, FALSE, FALS~
$ NatlHoliday  <lgl> TRUE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA~
$ TV_Coverage  <chr> "ESPN", "ACCNX", "ACCN", "ABC", "ACCN", "ESPN", "ACCN", "~
$ City         <chr> "Durham", "Durham", "Durham", "Durham", "Durham", "Durham~
$ State        <chr> "NC", "NC", "NC", "NC", "NC", "NC", "NC", "NC", "NC", "NC~
$ Bowl         <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, F~
$ UNC_Game     <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, F~
$ Type         <chr> "Rushing", "Rushing", "Rushing", "Rushing", "Rushing", "R~
$ Attempts     <dbl> 30, 45, 40, 40, 30, 41, 30, 35, 35, 48, 42, 41, 30, 34, 2~
$ Yards        <dbl> 199, 261, 268, 189, 194, 181, 69, 172, 222, 248, 297, 165~
$ AvgYd        <dbl> 6.633333, 5.800000, 6.700000, 4.725000, 6.466667, 4.41463~
$ TD_Gained    <dbl> 3, 4, 5, 1, 1, 2, 1, 1, 4, 4, 4, 1, 0, 0, 2, 0, 1, 2, 1, ~
$ Comp         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 17, 2~
$ CompPct      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 50.00~
$ Int          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 0, 0,~
$ Rating       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 93.23~
$ Touchbacks   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
$ TouchbackPct <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
$ OutOfBounds  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
$ Onside       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
$ Fumbles      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
$ isHome       <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRU~
```

Not all columns are used for each type of offensive statistic. For example, the "Onside" column is only relevant for rows whose *Type* column value is "Kickoffs". The *Comp* column represents completions (in terms of completed passes), successes (with Field_Goals, 3rd_Down_Conv,

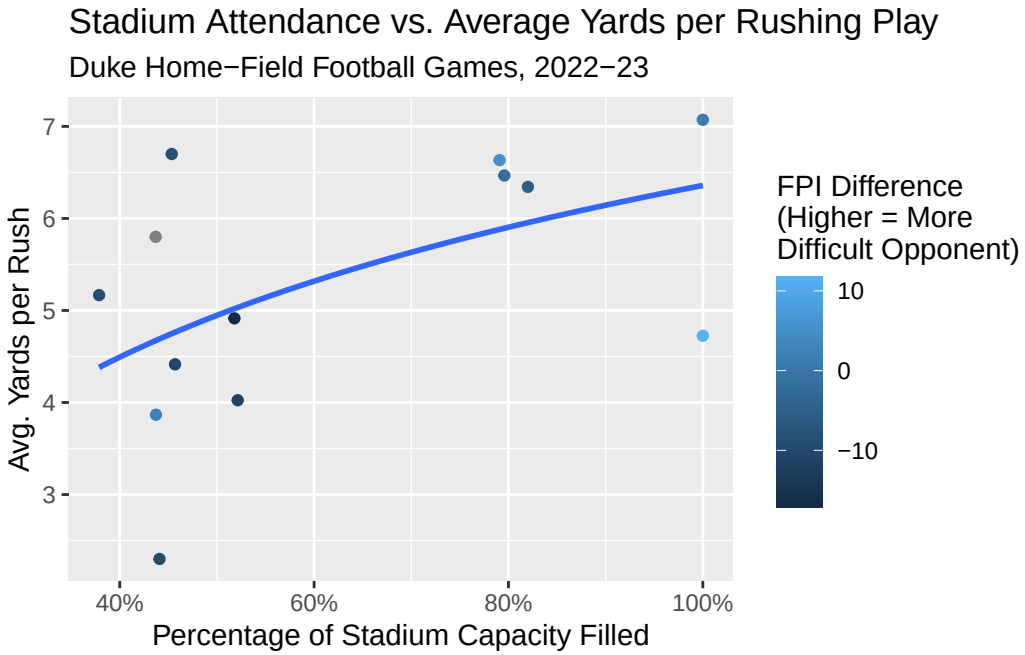
4th_Down_Conv, etc.), or a total count (with Duke_Penalties, Opp_Penalties, etc.) depending on the football context of the row's *Type*.

Rushing

Attendance as a predictor of *average yards* gained/lost per rushing play:

```
# Dataset filtering
home_off_rush_data <- home_offense_data |>
  filter(Type == "Rushing")

# Visualization
home_off_rush_data |>
  ggplot(
    aes(x = AttPct, y = AvgYd, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Average Yards per Rushing Play",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Avg. Yards per Rush",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_rush_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(AvgYd ~ log(AttPct), data = home_off_rush_data)

tidy(att_rush_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    -3.02      4.24    -0.712  0.492
2 log(AttPct)     2.04      1.04     1.96   0.0760
```

```
glance(att_rush_glm)$AIC
```

```
[1] 46.70963
```

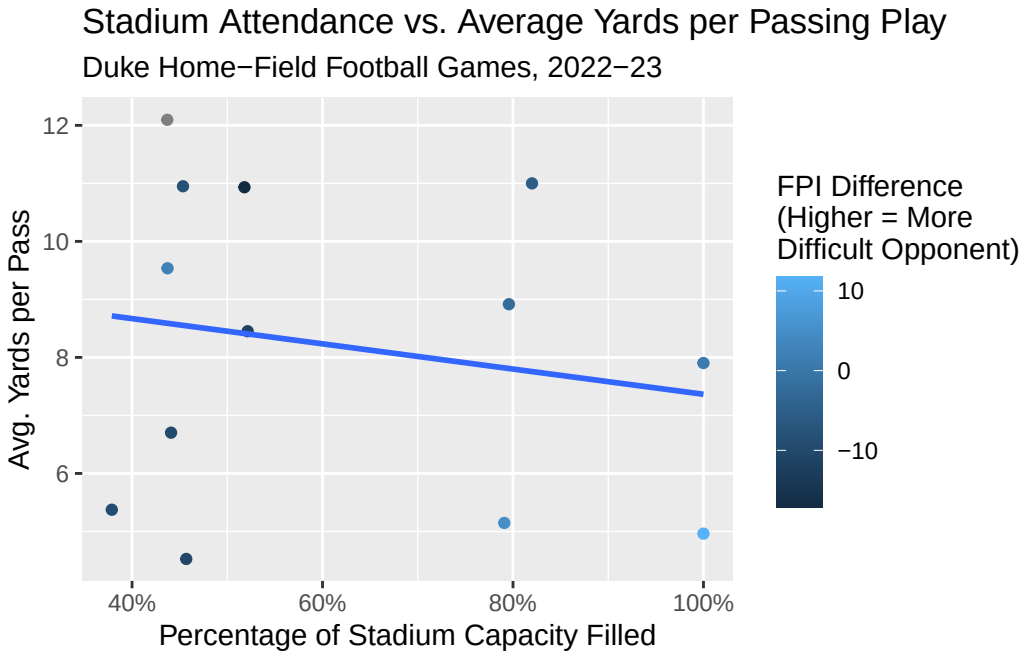
Wallace Wade attendance was *not* a strongly significant predictor of average yards gained/lost per rushing play in 2022-23.

Passing

Attendance as a predictor of *average yards* gained/lost per passing play:

```
# Dataset filtering
home_off_pass_data <- home_offense_data |>
  filter(Type == "Passing")

# Visualization
home_off_pass_data |>
  ggplot(
    aes(x = AttPct, y = AvgYd, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ x, se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Average Yards per Passing Play",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Avg. Yards per Pass",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_pass_yd_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(AvgYd ~ AttPct, data = home_off_pass_data)

tidy(att_pass_yd_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   9.54      2.26       4.21  0.00145
2 AttPct       -0.0217    0.0345    -0.631 0.541
```

```
glance(att_pass_yd_glm)$AIC
```

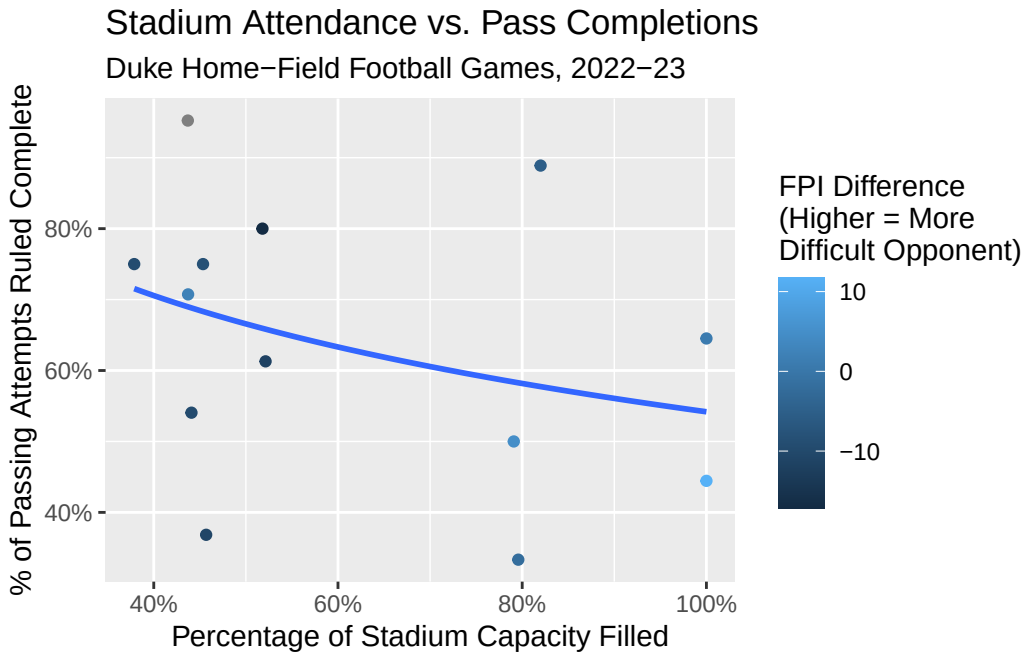
```
[1] 66.68931
```

Wallace Wade attendance was *not* a statistically significant predictor of average yards gained/lost per passing play in 2022-23.

Attendance as a predictor of passing *completions* per game:

```
# Visualization
home_off_pass_data |>
  ggplot(
    aes(x = AttPct, y = CompPct, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  scale_y_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Pass Completions",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "% of Passing Attempts Ruled Complete",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```



```r
# Linear model
att_pass_comp_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(CompPct ~ log(AttPct), data = home_off_pass_data)

tidy(att_pass_comp_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    137.       64.5      2.12  0.0580
2 log(AttPct)    -17.9      15.8     -1.13  0.282
```

```r
glance(att_pass_comp_glm)$AIC
```

```
[1] 117.4746
```

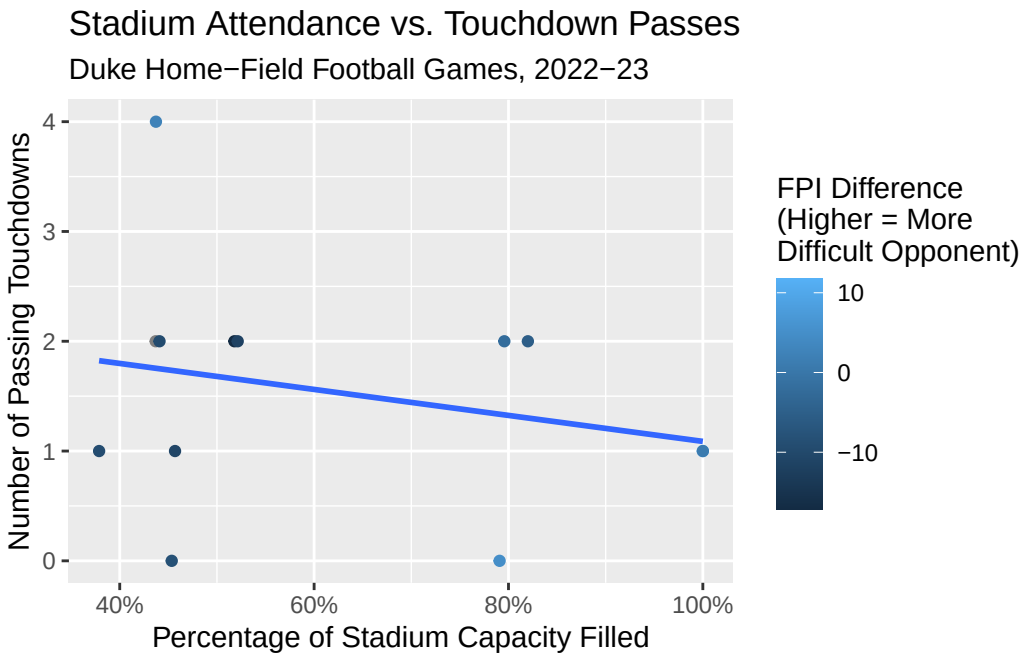
Wallace Wade attendance was *not* a statistically significant predictor of the percentage of passing plays that were completed per game in 2022-23.

Attendance as a predictor of *touchdown* passes per game:

```r
# Visualization
home_off_pass_data |>
  ggplot(
    aes(x = AttPct, y = TD_Gained, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ x, se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Touchdown Passes",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Passing Touchdowns",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
```

```
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_pass_td_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(TD_Gained ~ AttPct, data = home_off_pass_data)

tidy(att_pass_td_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   2.27      0.884      2.57   0.0260
2 AttPct       -0.0118    0.0135    -0.879  0.398
```

```
glance(att_pass_td_glm)$AIC
```

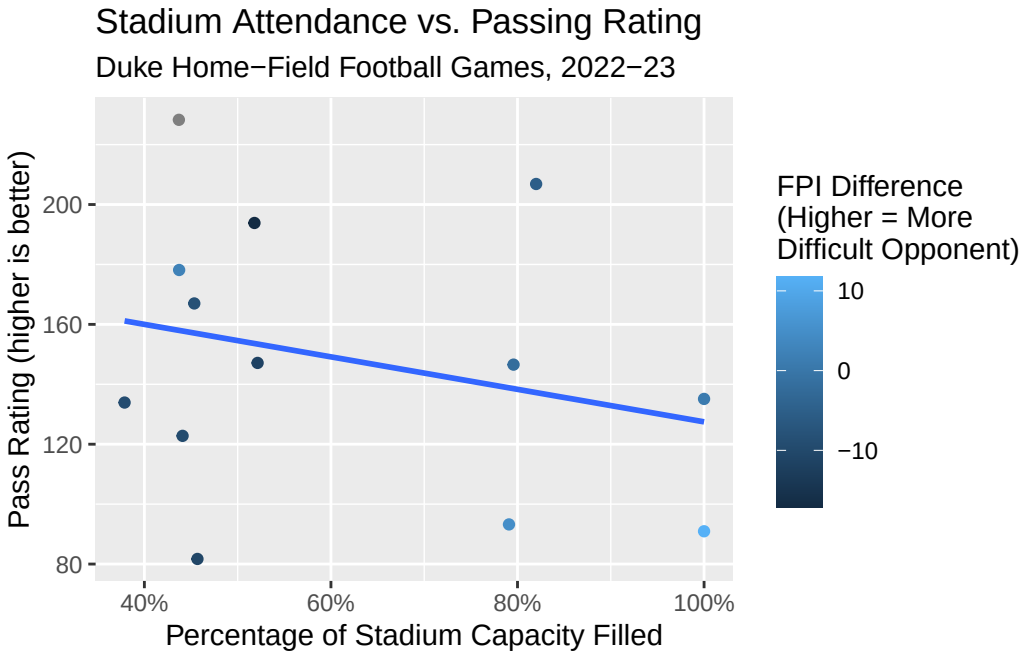
```
[1] 42.2379
```

Wallace Wade attendance was *not* a statistically significant predictor of the number of touchdown passing plays per game in 2022-23.

Attendance as a predictor of *pass rating*:

```
# Visualization
home_off_pass_data |>
  ggplot(
    aes(x = AttPct, y = Rating, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ x, se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Passing Rating",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Pass Rating (higher is better)",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_pass_qb_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Rating ~ AttPct, data = home_off_pass_data)

tidy(att_pass_qb_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  182.       38.2       4.76  0.000595
2 AttPct        -0.542     0.582    -0.932 0.371
```

```
glance(att_pass_qb_glm)$AIC
```

```
[1] 140.1709
```

Wallace Wade attendance was *not* a statistically significant predictor of passing rating per game in 2022-23.

Punt Returns

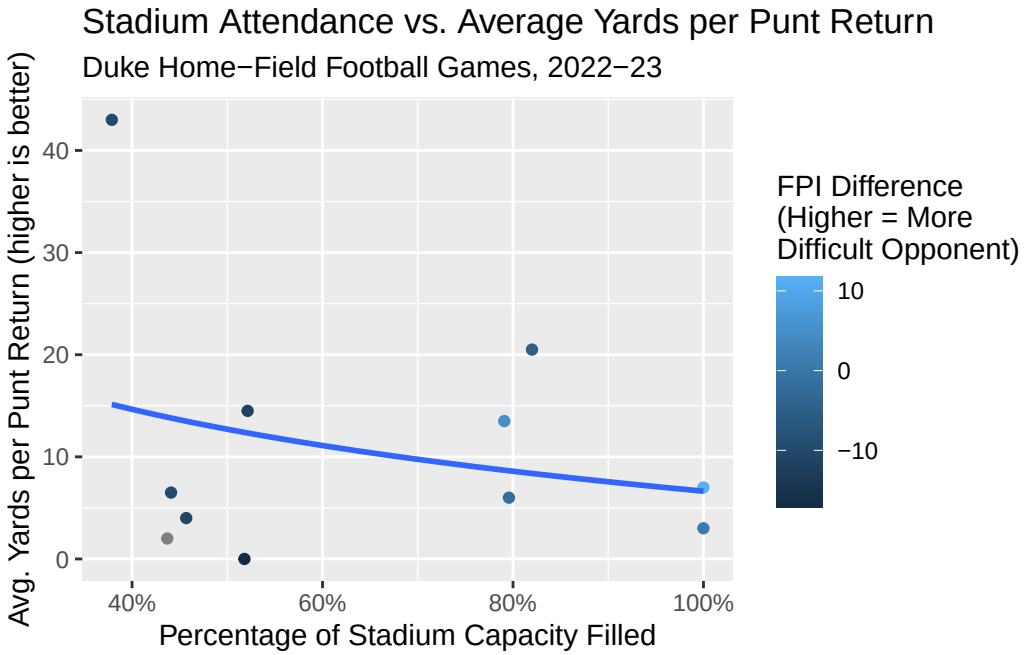
Attendance as a predictor of *average yards* returned per punt return:

```
# Dataset filtering
home_off_punt_return_data <- home_offense_data |>
  filter(Type == "Punt_Returns")

# Visualization
home_off_punt_return_data |>
  ggplot(
    aes(x = AttPct, y = AvgYd, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Average Yards per Punt Return",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Avg. Yards per Punt Return (higher is better)",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```

```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```




```
# Linear model
att_punt_ret_yd_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(AvgYd ~ log(AttPct), data = home_off_punt_return_data)

tidy(att_punt_ret_yd_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    46.9       45.5     1.03    0.329
2 log(AttPct)    -8.75      11.0    -0.795   0.447
```

```
glance(att_punt_ret_yd_glm)$AIC
```

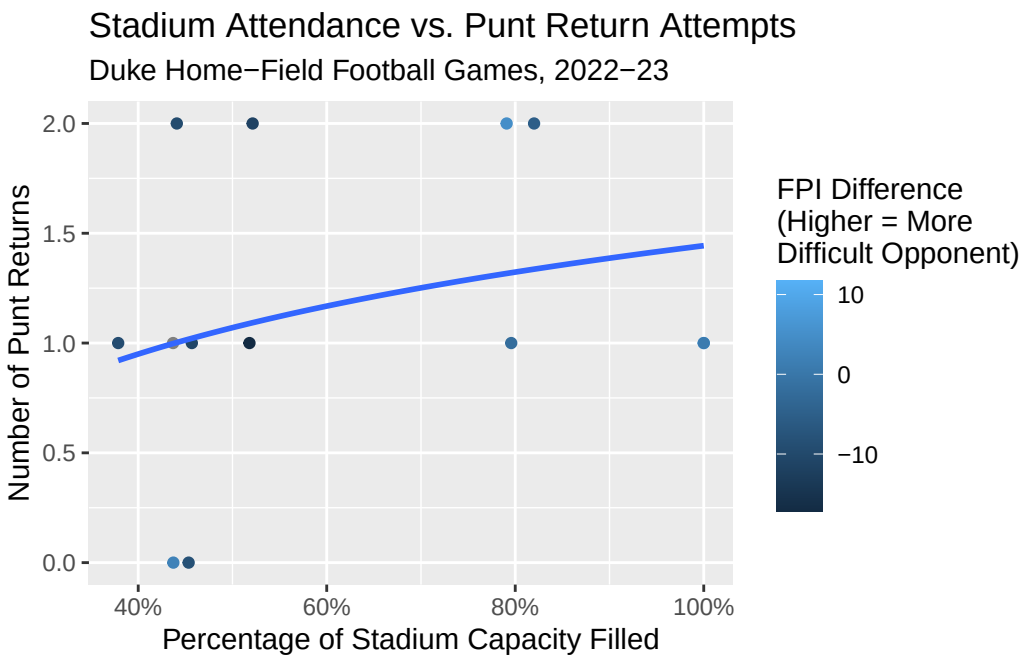
```
[1] 90.60412
```

Wallace Wade attendance was *not* a statistically significant predictor of average punt return yardage per game in 2022-23.

Attendance as a predictor of punt return *attempts* per game:

```
# Visualization
home_off_punt_return_data |>
  ggplot(
    aes(x = AttPct, y = Attempts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Punt Return Attempts",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Punt Returns",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```



```r
# Linear model
att_punt_ret_attempts_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Attempts ~ log(AttPct), data = home_off_punt_return_data)

tidy(att_punt_ret_attempts_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   -1.04      2.33     -0.445   0.665
2 log(AttPct)    0.539     0.571     0.943   0.366
```

```r
glance(att_punt_ret_attempts_glm)$AIC
```

```
[1] 31.14664
```

Wallace Wade attendance was *not* a statistically significant predictor of the number of punt return attempts per game in 2022-23.

Kickoff Returns

Attendance as a predictor of *average yards* returned per kickoff return:

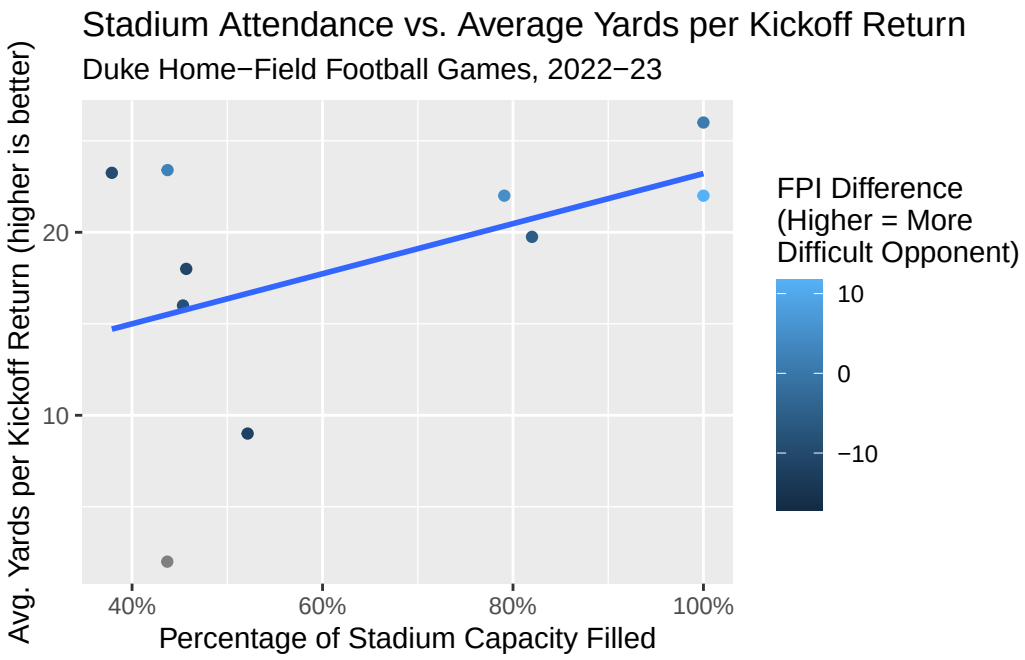
```r
# Dataset filtering
home_off_kickoff_return_data <- home_offense_data |>
  filter(Type == "Kickoff_Returns")

# Visualization
home_off_kickoff_return_data |>
  ggplot(
    aes(x = AttPct, y = AvgYd, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ x, se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Average Yards per Kickoff Return",
       subtitle = "Duke Home-Field Football Games, 2022-23",
```

```
        x = "Percentage of Stadium Capacity Filled",
        y = "Avg. Yards per Kickoff Return (higher is better)",
        color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```

```
Warning: Removed 3 rows containing missing values (`geom_point()`).
```




```
# Linear model
att_kickoff_ret_yd_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(AvgYd ~ AttPct, data = home_off_kickoff_return_data)

tidy(att_kickoff_ret_yd_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    9.52      6.38       1.49   0.174
2 AttPct         0.137     0.0950     1.44   0.188
```

```
glance(att_kickoff_ret_yd_glm)$AIC
```

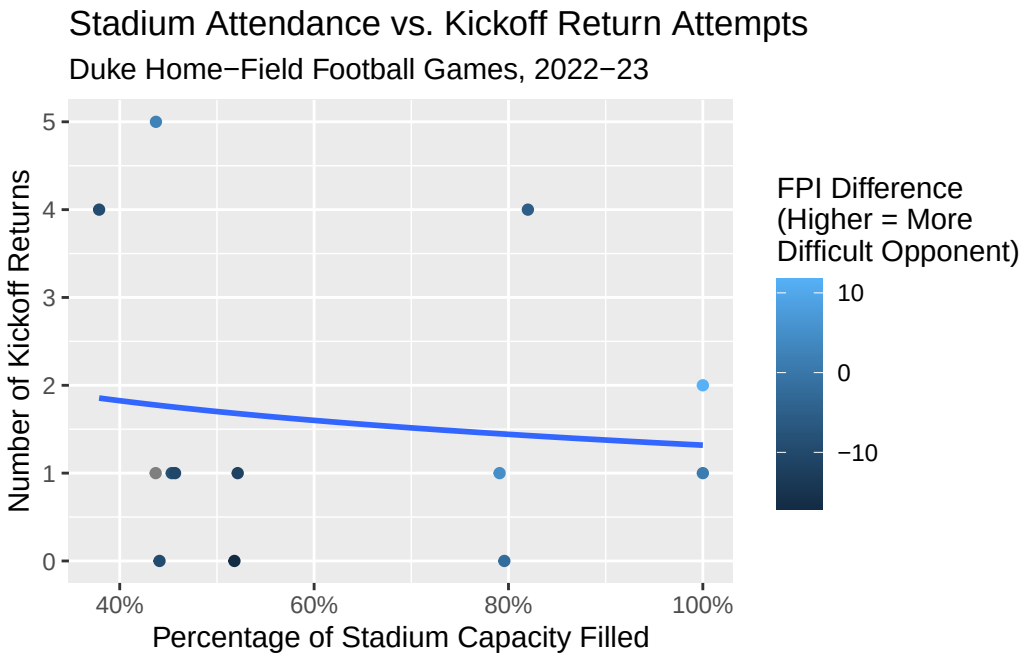
```
[1] 71.1234
```

Wallace Wade attendance was *not* a statistically significant predictor of the average yards returned per kickoff return attempt in 2022-23.

Attendance as a predictor of kickoff return *attempts* per game:

```
# Visualization
home_off_kickoff_return_data |>
  ggplot(
    aes(x = AttPct, y = Attempts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Kickoff Return Attempts",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Kickoff Returns",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_kickoff_ret_attempts_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Attempts ~ log(AttPct), data = home_off_kickoff_return_data)

tidy(att_kickoff_ret_attempts_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    3.86       5.80     0.666   0.519
2 log(AttPct)   -0.552      1.42    -0.389   0.705
```

```
glance(att_kickoff_ret_attempts_glm)$AIC
```

```
[1] 54.85574
```

Wallace Wade attendance was *not* a statistically significant predictor of the number of kickoff return attempts per game in 2022-23.

Punts

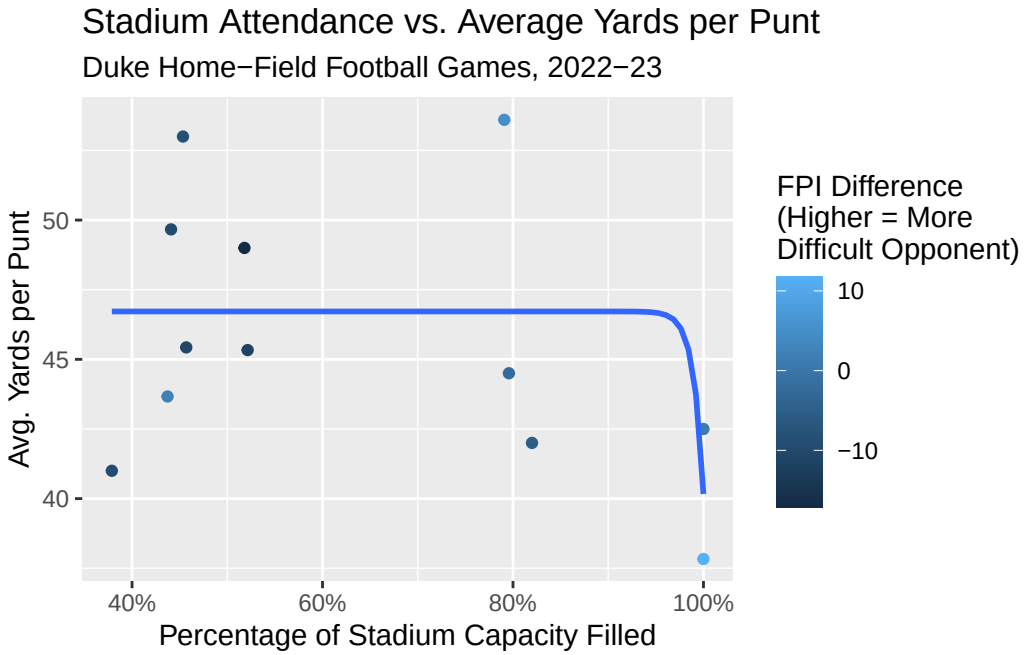
Attendance as a predictor of *average yards* per punt:

```
# Dataset filtering
home_off_punts_data <- home_offense_data |>
  filter(Type == "Punts")

# Visualization
home_off_punts_data |>
  ggplot(
    aes(x = AttPct, y = AvgYd, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ exp(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Average Yards per Punt",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Avg. Yards per Punt",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```

```
Warning: Removed 1 rows containing missing values (`geom_point()`).
```




```
# Linear model
att_punts_yd_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(AvgYd ~ exp(AttPct), data = home_off_punts_data)

tidy(att_punts_yd_glm)
```

```
# A tibble: 2 x 5
  term         estimate std.error statistic  p.value
  <chr>           <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  4.67e+ 1  1.36e+ 0     34.4  1.02e-11
2 exp(AttPct) -2.44e-43  1.24e-43     -1.97 7.71e- 2
```

```
glance(att_punts_yd_glm)$AIC
```

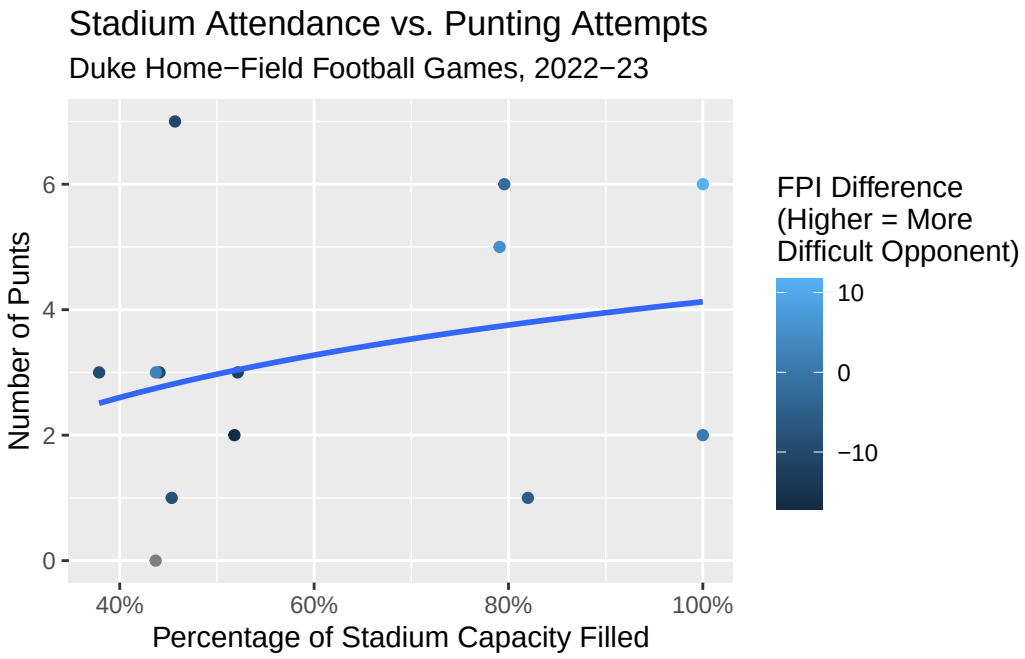
```
[1] 72.83974
```

Wallace Wade attendance was *not* a strong predictor of the average yards punted per punt attempt in 2022-23.

Attendance as a predictor of punting *attempts* per game:

```
# Visualization
home_off_punts_data |>
  ggplot(
    aes(x = AttPct, y = Attempts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Punting Attempts",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Punts",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```



```r
# Linear model
att_punts_attempts_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Attempts ~ log(AttPct), data = home_off_punts_data)

tidy(att_punts_attempts_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    -3.55      7.34    -0.483   0.638
2 log(AttPct)     1.67      1.80     0.926   0.374
```

```r
glance(att_punts_attempts_glm)$AIC
```

```
[1] 60.97246
```

Wallace Wade attendance was *not* a statistically significant predictor of the number of punting attempts per game in 2022-23.

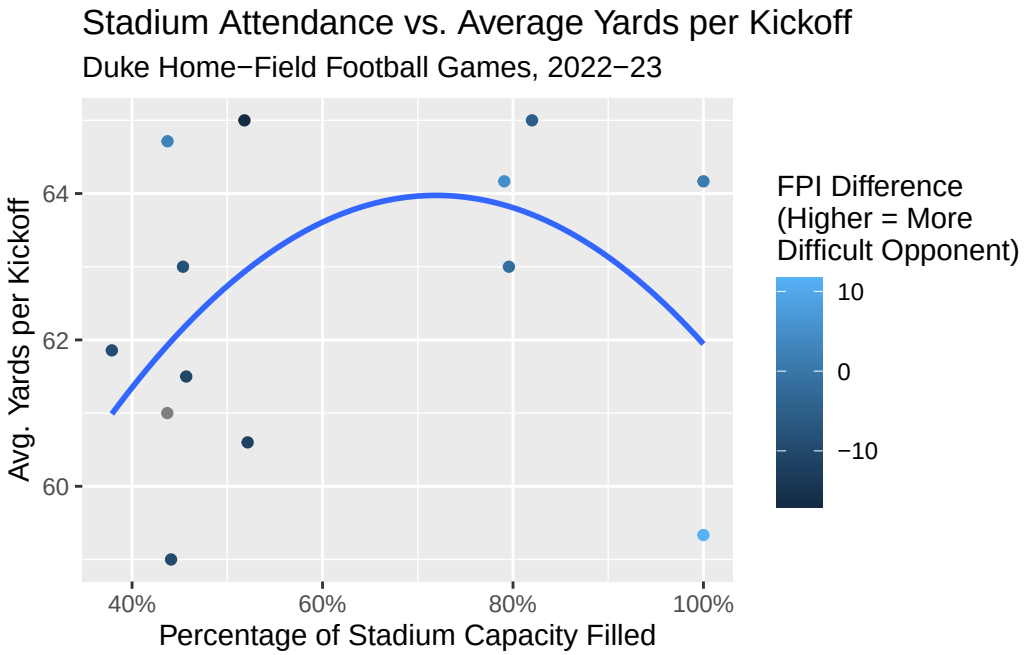
Kickoffs

Attendance as a predictor of *average yards* per kickoff:

```r
# Dataset filtering
home_off_kickoffs_data <- home_offense_data |>
  filter(Type == "Kickoffs")

# Visualization
home_off_kickoffs_data |>
  ggplot(
    aes(x = AttPct, y = AvgYd, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ poly(x,2), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Average Yards per Kickoff",
       subtitle = "Duke Home-Field Football Games, 2022-23",
```

```
        x = "Percentage of Stadium Capacity Filled",
        y = "Avg. Yards per Kickoff",
        color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_kickoffs_yd_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(AvgYd ~ poly(AttPct,2), data = home_off_kickoffs_data)

tidy(att_kickoffs_yd_glm)
```

```
# A tibble: 3 x 5
  term             estimate std.error statistic  p.value
  <chr>               <dbl>     <dbl>     <dbl>    <dbl>
```

```
1 (Intercept)         62.5      0.579   108.    1.14e-16
2 poly(AttPct, 2)1     1.23     2.09      0.588 5.70e- 1
3 poly(AttPct, 2)2    -2.82     2.09     -1.35  2.06e- 1
```

```
glance(att_kickoffs_yd_glm)$AIC
```

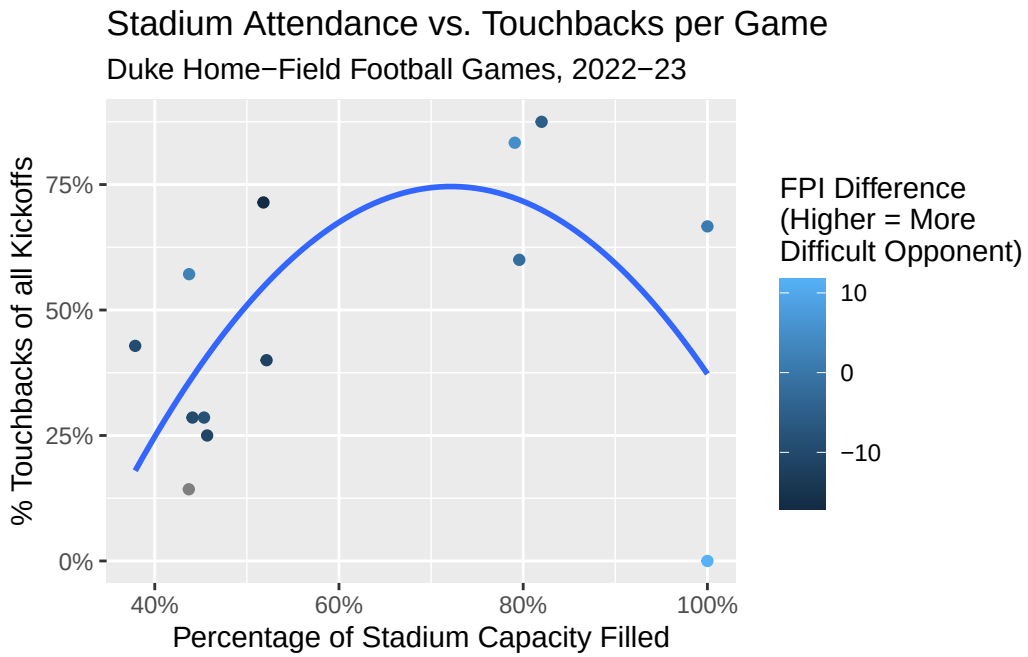
```
[1] 60.60392
```

Wallace Wade attendance was *not* a strong predictor of the average yards kicked per kickoff attempt in 2022-23.

Attendance as a predictor of *touchback percentage* per game:

```
# Visualization
home_off_kickoffs_data |>
  ggplot(
    aes(x = AttPct, y = TouchbackPct, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ poly(x,2), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  scale_y_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Touchbacks per Game",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "% Touchbacks of all Kickoffs",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_kickoffs_touchback_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(TouchbackPct ~ poly(AttPct,2), data = home_off_kickoffs_data)

tidy(att_kickoffs_touchback_glm)
```

```
# A tibble: 3 x 5
  term             estimate std.error statistic   p.value
  <chr>               <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)          46.6      6.40      7.28 0.0000267
2 poly(AttPct, 2)1     24.7     23.1       1.07 0.309
3 poly(AttPct, 2)2    -52.8     23.1      -2.29 0.0451
```

```
glance(att_kickoffs_touchback_glm)$AIC
```

```
[1] 123.0841
```

Wallace Wade attendance is a statistically ***possible*** predictor of the percent of kickoff attempts that were touchbacks in 2022-23.

Field Goals

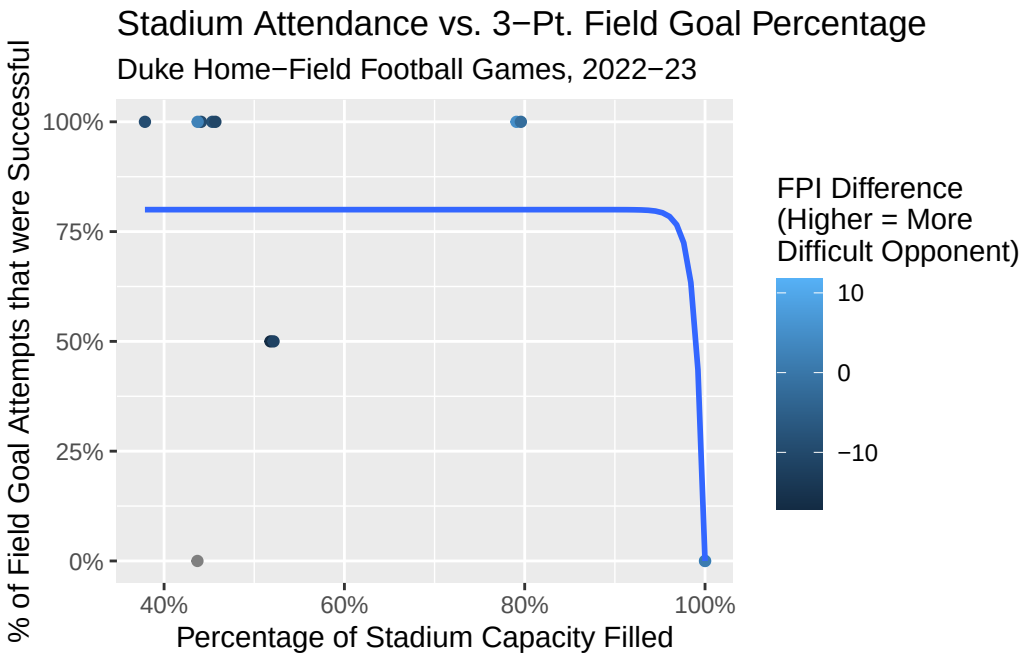
Attendance as a predictor of 3-point field goal percentage per game:

```
# Dataset filtering
home_off_fg_data <- home_offense_data |>
  filter(Type == "Field_Goals")

# Visualization
home_off_fg_data |>
  ggplot(
    aes(x = AttPct, y = CompPct, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ exp(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  scale_y_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. 3-Pt. Field Goal Percentage",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "% of Field Goal Attempts that were Successful",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```

```
Warning: Removed 1 rows containing missing values (`geom_point()`).
```




```
# Linear model
att_fg_comp_pct_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(CompPct ~ exp(AttPct), data = home_off_fg_data)

tidy(att_fg_comp_pct_glm)
```

```
# A tibble: 2 x 5
  term          estimate std.error statistic   p.value
  <chr>            <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)  8.00e+ 1  1.05e+ 1      7.63 0.0000178
2 exp(AttPct) -2.98e-42  9.56e-43     -3.11 0.0110
```

```
glance(att_fg_comp_pct_glm)$AIC
```

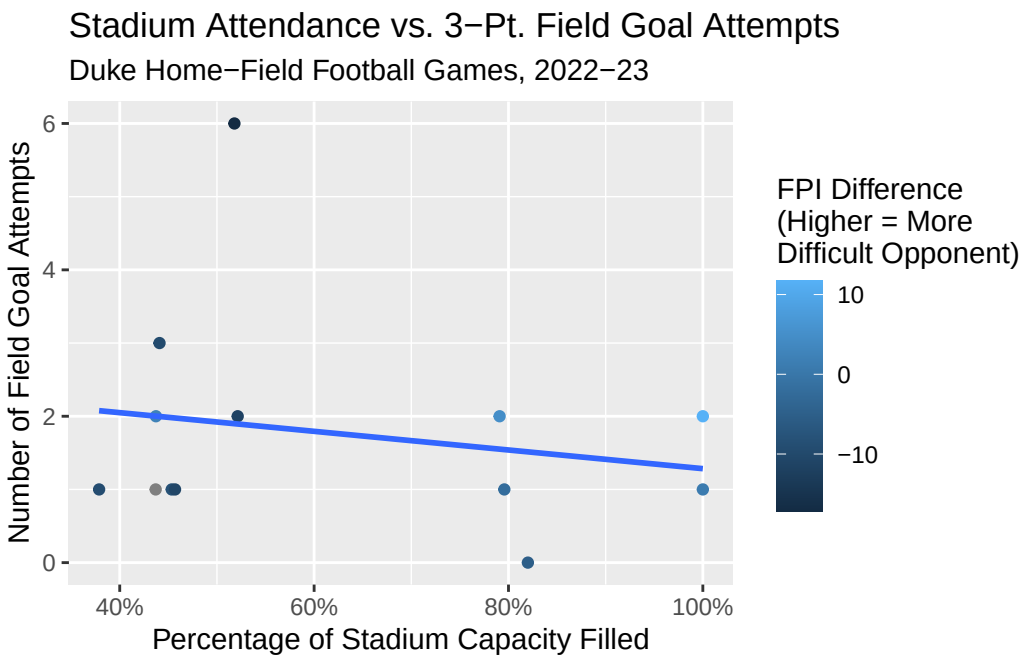
```
[1] 121.9035
```

Based on Duke football games in Wallace Wade during 2022-23, there ***is*** significant evidence to suggest that when football attendance is 100% or very close to 100%, the predicted 3-pt. field goals percentage decreases dramatically.

Attendance as a predictor of the number of 3-pt. field goals *attempted* per game:

```
# Visualization
home_off_fg_data |>
  ggplot(
    aes(x = AttPct, y = Attempts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ x, se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. 3-Pt. Field Goal Attempts",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Field Goal Attempts",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```



```r
# Linear model
att_fg_attempts_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Attempts ~ AttPct, data = home_off_fg_data)

tidy(att_fg_attempts_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   2.56      1.26       2.02   0.0680
2 AttPct       -0.0127    0.0193    -0.662  0.521
```

```r
glance(att_fg_attempts_glm)$AIC
```

```
[1] 51.54802
```

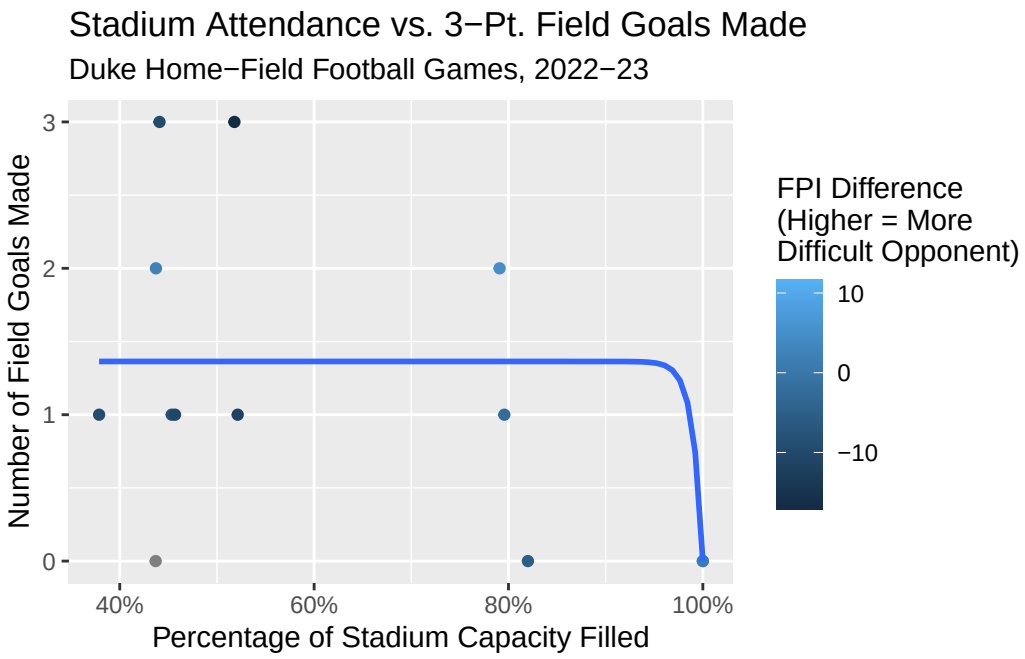
Wallace Wade attendance was *not* a statistically significant predictor of the number of 3-point field goals attempted per game in 2022-23.

Attendance as a predictor of the number of 3-pt. field goals *made* per game:

```r
# Visualization
home_off_fg_data |>
  ggplot(
    aes(x = AttPct, y = Comp, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ exp(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. 3-Pt. Field Goals Made",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Field Goals Made",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
```

```
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_fg_comp_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Comp ~ exp(AttPct), data = home_off_fg_data)

tidy(att_fg_comp_glm)
```

```
# A tibble: 2 x 5
  term         estimate std.error statistic  p.value
  <chr>           <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  1.36e+ 0  2.95e- 1      4.62 0.000741
2 exp(AttPct) -5.07e-44  2.80e-44     -1.81 0.0974
```

```
glance(att_fg_comp_glm)$AIC
```

```
[1] 40.17209
```

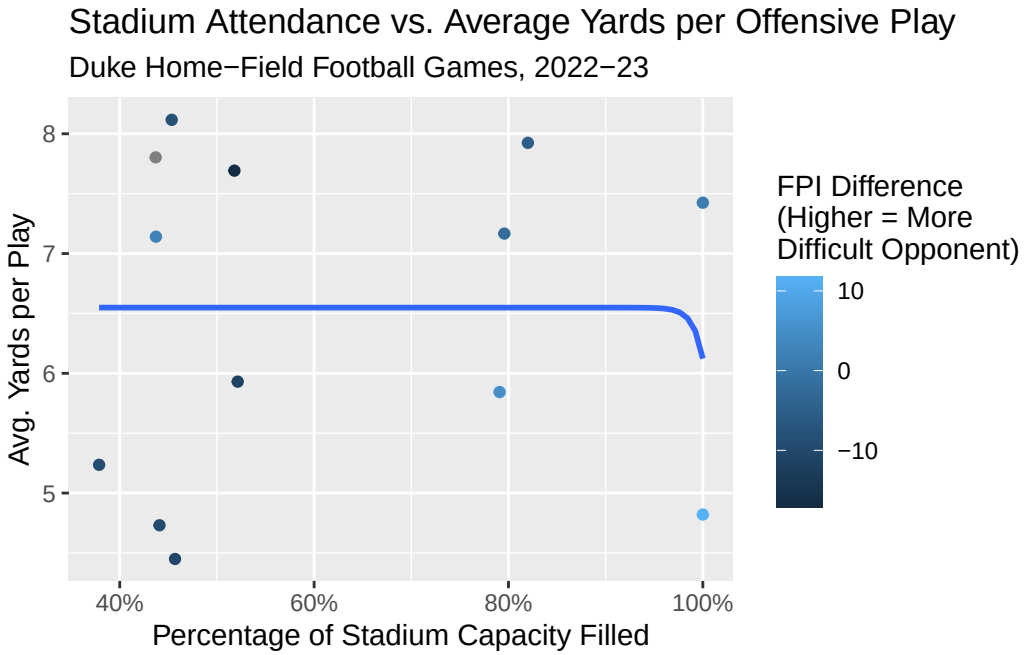
Wallace Wade attendance was *not* a *strong* predictor of the number of 3-point field goals made per game in 2022-23. (However, there may be a slightly stronger, albiet still weak, association between FGs made & stadium attendance than the association between FGs attempted & stadium attendance.)

Total Offense

Attendance as a predictor of *average yards* per offensive play:

```
# Dataset filtering
home_off_total_data <- home_offense_data |>
  filter(Type == "Total_Offense")

# Visualization
home_off_total_data |>
  ggplot(
    aes(x = AttPct, y = AvgYd, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ exp(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Average Yards per Offensive Play",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Avg. Yards per Play",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_total_yd_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(AvgYd ~ exp(AttPct), data = home_off_total_data)

tidy(att_total_yd_glm)
```

```
# A tibble: 2 x 5
  term         estimate std.error statistic       p.value
  <chr>           <dbl>     <dbl>     <dbl>         <dbl>
1 (Intercept)  6.55e+ 0  4.23e- 1    15.5   0.00000000820
2 exp(AttPct) -1.58e-44  4.01e-44    -0.395 0.701
```

```
glance(att_total_yd_glm)$AIC
```

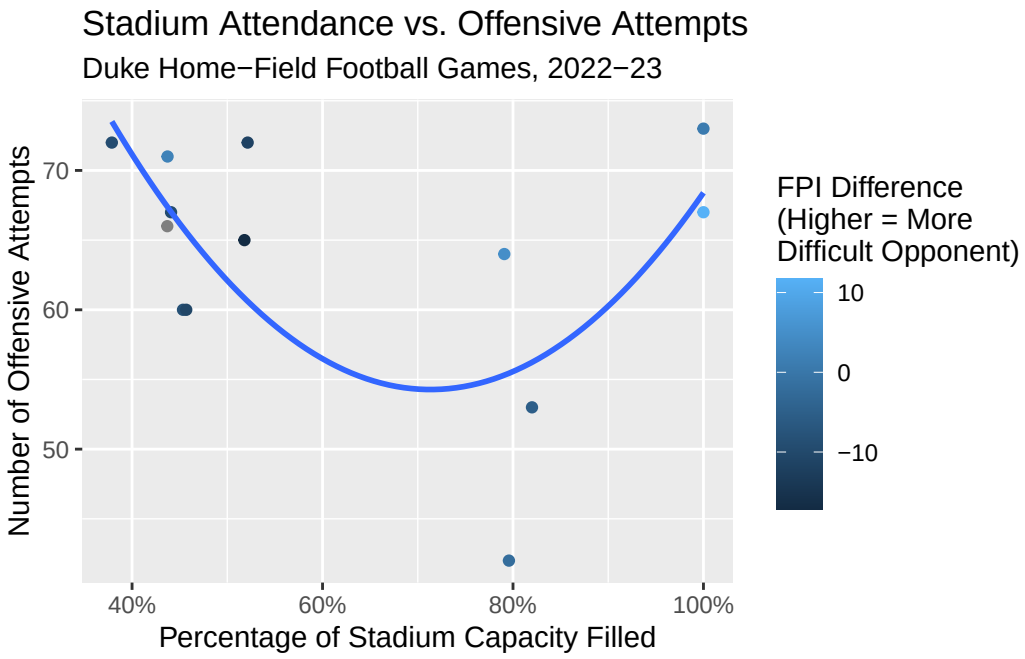
```
[1] 49.53165
```

Wallace Wade attendance was *not* a statistically significant predictor of the average yards gained/lost per offensive play attempt in 2022-23.

Attendance as a predictor of total offensive *attempts* per game:

```
# Visualization
home_off_total_data |>
  ggplot(
    aes(x = AttPct, y = Attempts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ poly(x,2), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Offensive Attempts",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Offensive Attempts",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```



```r
# Linear model
att_total_attempts_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Attempts ~ poly(AttPct,2), data = home_off_total_data)

tidy(att_total_attempts_glm)
```

```
# A tibble: 3 x 5
  term              estimate std.error statistic  p.value
  <chr>                <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)          64         2.00    32.0   2.09e-11
2 poly(AttPct, 2)1     -6.59      7.21    -0.914 3.82e- 1
3 poly(AttPct, 2)2     18.8       7.21     2.61  2.60e- 2
```

```r
glance(att_total_attempts_glm)$AIC
```

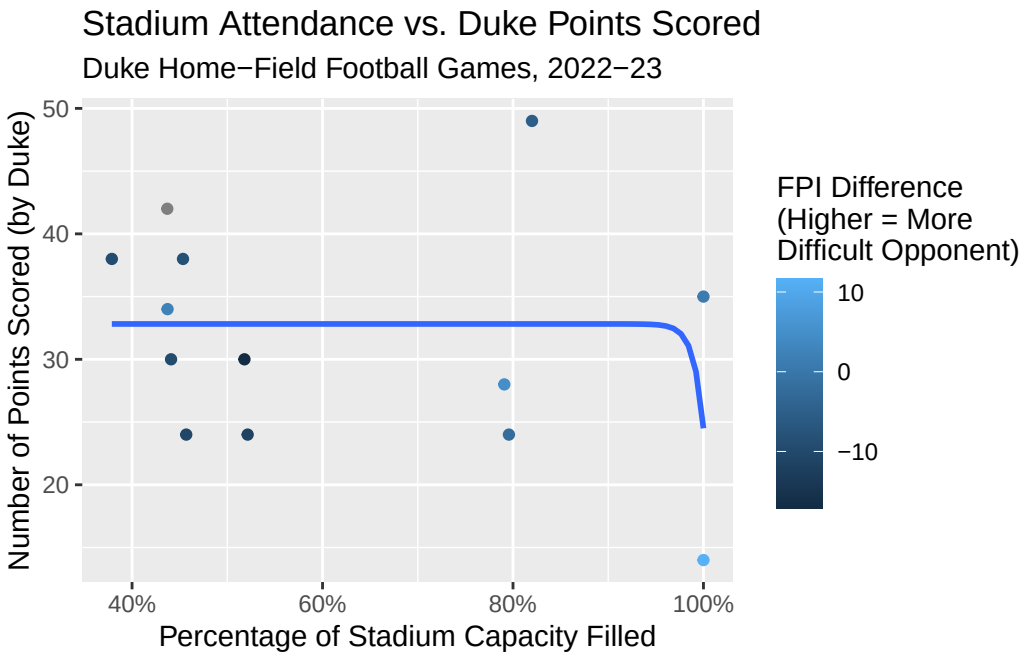
```
[1] 92.84864
```

Wallace Wade attendance was *not* a statistically significant predictor of the number of offensive attempts per game in 2022-23.

Attendance as a predictor of total *points scored* by *Duke* per game:

```r
# Visualization
home_off_total_data |>
  ggplot(
    aes(x = AttPct, y = DukePts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ exp(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Duke Points Scored",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Points Scored (by Duke)",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_total_pts_duke_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(DukePts ~ exp(AttPct), data = home_off_total_data)

tidy(att_total_pts_duke_glm)
```

```
# A tibble: 2 x 5
  term         estimate std.error statistic     p.value
  <chr>           <dbl>     <dbl>     <dbl>       <dbl>
1 (Intercept)  3.28e+ 1  2.72e+ 0     12.1  0.000000109
2 exp(AttPct) -3.09e-43  2.58e-43     -1.20 0.255
```

```
glance(att_total_pts_duke_glm)$AIC
```

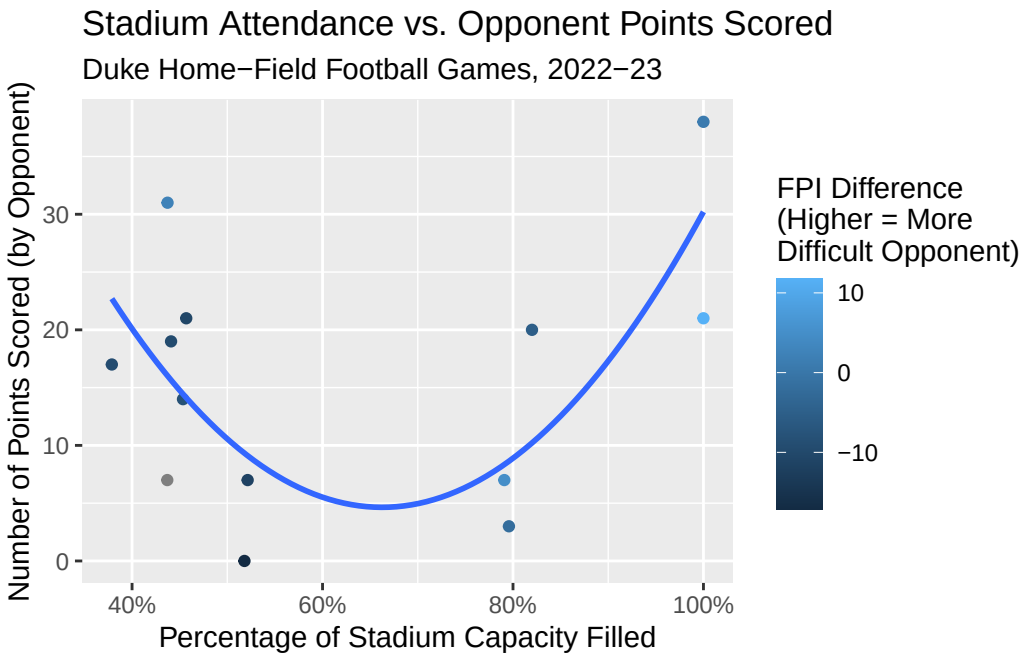
```
[1] 97.89422
```

Wallace Wade attendance was *not* a statistically significant predictor of the number of points scored by Duke per game in 2022-23.

Attendance as a predictor of total *points scored* by Duke's *opponent* per game:

```
# Visualization
home_off_total_data |>
  ggplot(
    aes(x = AttPct, y = OppPts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ poly(x,2), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Opponent Points Scored",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Points Scored (by Opponent)",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_total_pts_opp_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(OppPts ~ poly(AttPct,2), data = home_off_total_data)

tidy(att_total_pts_opp_glm)
```

```
# A tibble: 3 x 5
  term              estimate std.error statistic   p.value
  <chr>                <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)         15.8       2.42      6.51 0.0000684
2 poly(AttPct, 2)1     9.39      8.74      1.07 0.308
3 poly(AttPct, 2)2    24.6       8.74      2.81 0.0184
```

```
glance(att_total_pts_opp_glm)$AIC
```

```
[1] 97.84336
```

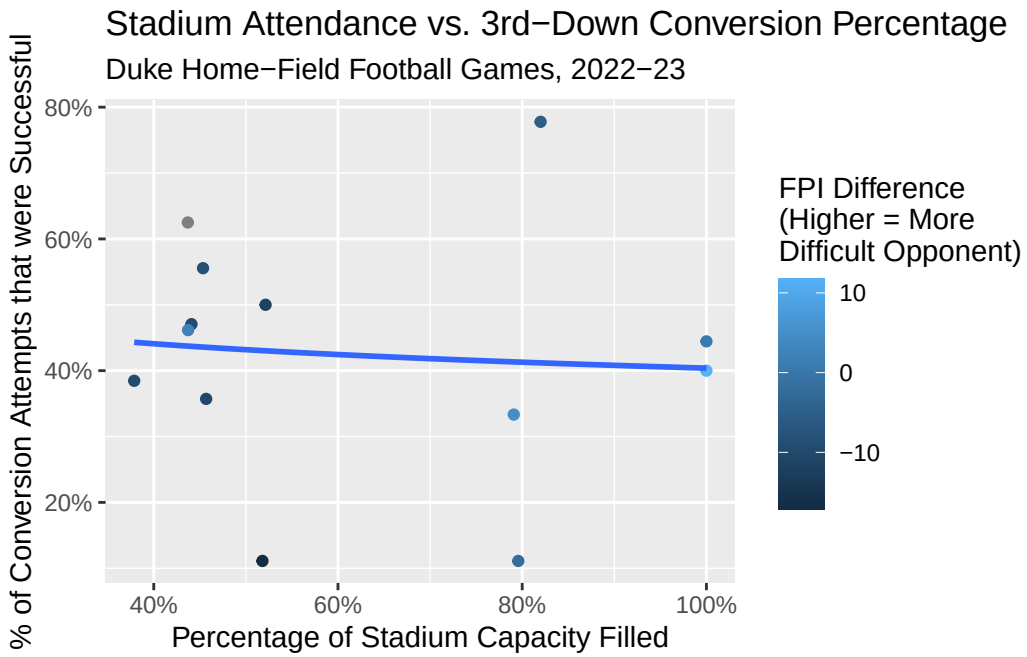
Wallace Wade attendance was *not* a statistically significant predictor of the number of points scored by Duke's opponent per game in 2022-23.

3rd Down Conversions

Attendance as a predictor of successful 3rd-down conversion *percentage* per game:

```
# Dataset filtering
home_off_3rd_conv_data <- home_offense_data |>
  filter(Type == "3rd_Down_Conv")

# Visualization
home_off_3rd_conv_data |>
  ggplot(
    aes(x = AttPct, y = CompPct, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  scale_y_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. 3rd-Down Conversion Percentage",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "% of Conversion Attempts that were Successful",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_3rd_conv_comp_pct_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(CompPct ~ log(AttPct), data = home_off_3rd_conv_data)

tidy(att_3rd_conv_comp_pct_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    59.0       64.3     0.917   0.379
2 log(AttPct)    -4.04      15.8    -0.256   0.802
```

```
glance(att_3rd_conv_comp_pct_glm)$AIC
```

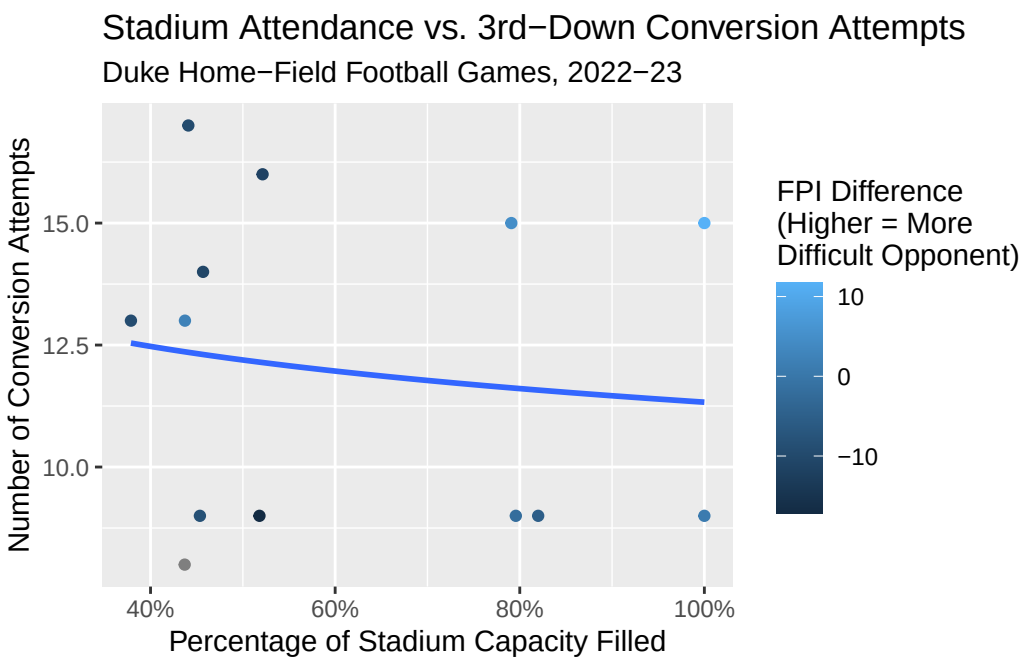
```
[1] 117.3976
```

Wallace Wade attendance was *not* a statistically significant predictor of the percentage of 3rd-down conversion attempts that were successful per game in 2022-23.

Attendance as a predictor of 3rd-down conversion *attempts* per game:

```
# Visualization
home_off_3rd_conv_data |>
  ggplot(
    aes(x = AttPct, y = Attempts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. 3rd-Down Conversion Attempts",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Conversion Attempts",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```



```r
# Linear model
att_3rd_conv_attempts_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Attempts ~ log(AttPct), data = home_off_3rd_conv_data)

tidy(att_3rd_conv_attempts_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    17.1      11.3       1.51    0.159
2 log(AttPct)    -1.25      2.77     -0.450   0.661
```

```r
glance(att_3rd_conv_attempts_glm)$AIC
```

```
[1] 72.18207
```

Wallace Wade attendance was *not* a statistically significant predictor of the number of 3rd-down conversion attempts per game in 2022-23.

4th Down Conversions

Attendance as a predictor of successful 4rd-down conversion *percentage* per game:

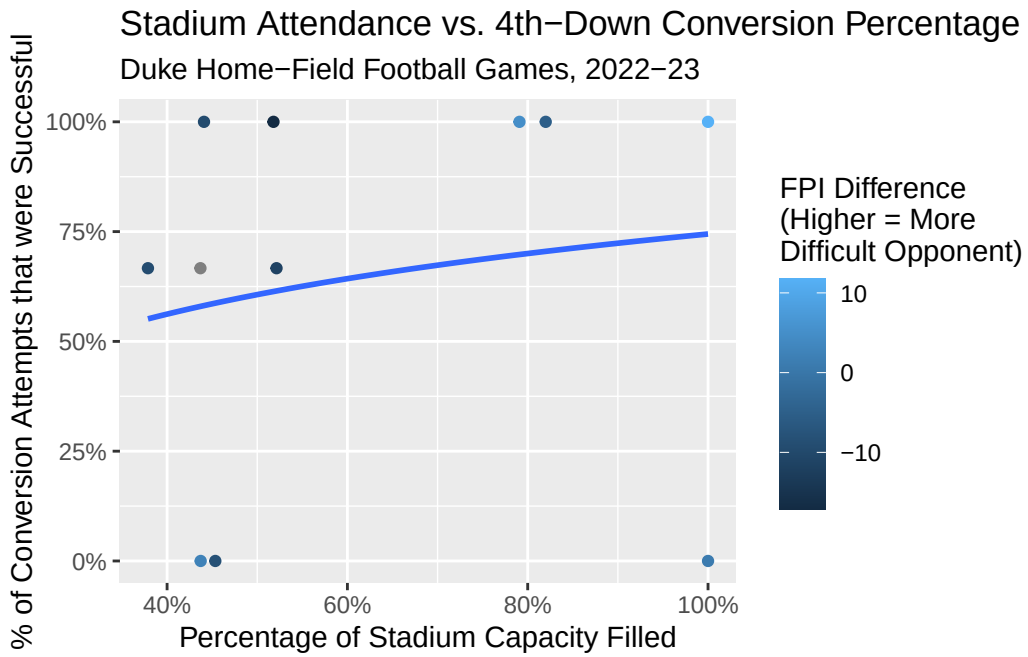
```r
# Dataset filtering
home_off_4th_conv_data <- home_offense_data |>
  filter(Type == "4th_Down_Conv")

# Visualization
home_off_4th_conv_data |>
  ggplot(
    aes(x = AttPct, y = CompPct, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  scale_y_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. 4th-Down Conversion Percentage",
```

```
        subtitle = "Duke Home-Field Football Games, 2022-23",
        x = "Percentage of Stadium Capacity Filled",
        y = "% of Conversion Attempts that were Successful",
        color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```

```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```




```
# Linear model
att_4th_comp_pct_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(CompPct ~ log(AttPct), data = home_off_4th_conv_data)
```

```
tidy(att_4th_comp_pct_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    -17.2     160.     -0.107   0.917
2 log(AttPct)     19.9      39.4     0.506   0.625
```

```
glance(att_4th_comp_pct_glm)$AIC
```

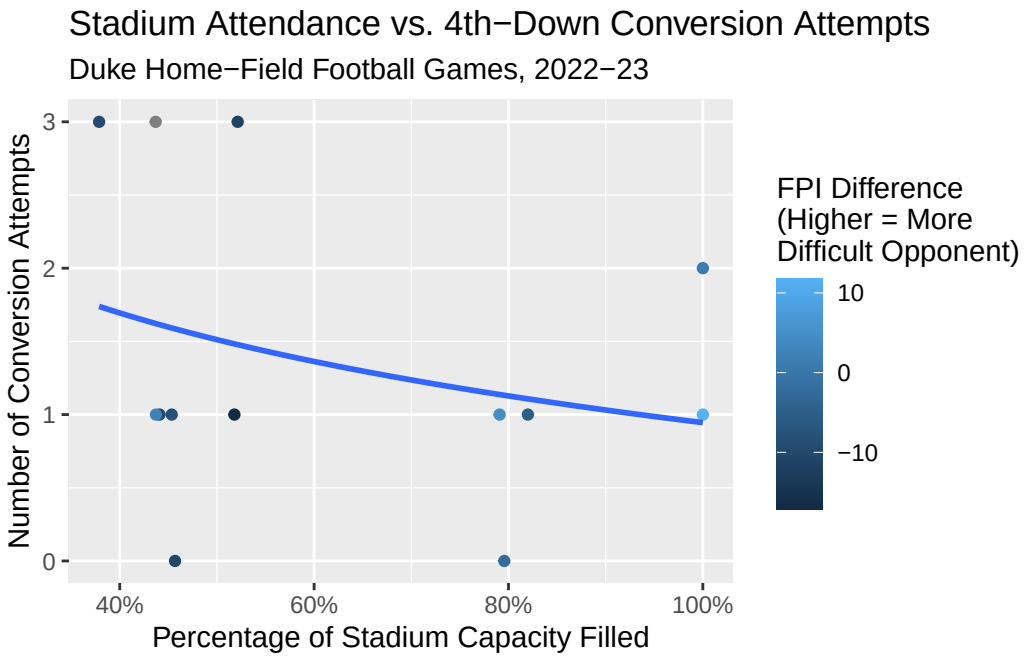
```
[1] 118.782
```

Wallace Wade attendance was *not* a statistically significant predictor of the percentage of 4th-down conversion attempts that were successful per game in 2022-23.

Attendance as a predictor of 4th-down conversion *attempts* per game:

```
# Visualization
home_off_4th_conv_data |>
  ggplot(
    aes(x = AttPct, y = Attempts, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. 4th-Down Conversion Attempts",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Conversion Attempts",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_4th_conv_attempts_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Attempts ~ log(AttPct), data = home_off_4th_conv_data)

tidy(att_4th_conv_attempts_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    4.71      3.53      1.33    0.209
2 log(AttPct)   -0.818     0.866    -0.945   0.365
```

```
glance(att_4th_conv_attempts_glm)$AIC
```

```
[1] 41.95507
```

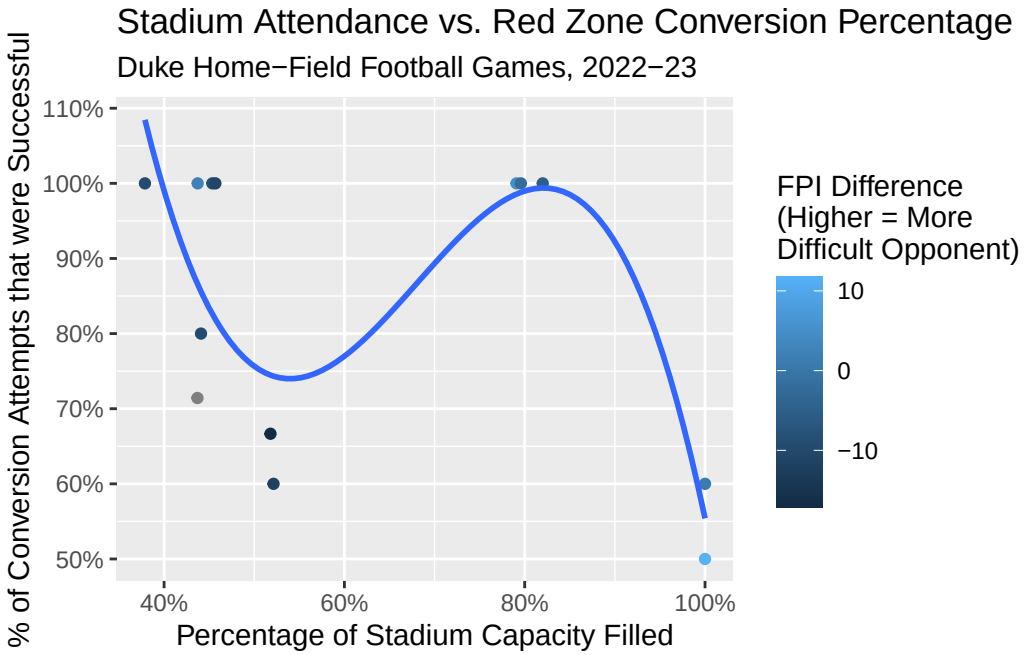
Wallace Wade attendance was *not* a statistically significant predictor of the number of 4th-down conversion attempts per game in 2022-23.

Red Zone Conversions

Attendance as a predictor of successful red zone conversion *percentage* per game:

```
# Dataset filtering
home_off_red_conv_data <- home_offense_data |>
  filter(Type == "Red_Zone_Conv")

# Visualization
home_off_red_conv_data |>
  ggplot(
    aes(x = AttPct, y = CompPct, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ poly(x,3), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  scale_y_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Red Zone Conversion Percentage",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "% of Conversion Attempts that were Successful",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_red_comp_pct_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(CompPct ~ poly(AttPct,3), data = home_off_red_conv_data)

tidy(att_red_comp_pct_glm)
```

```
# A tibble: 4 x 5
  term              estimate std.error statistic       p.value
  <chr>                <dbl>     <dbl>     <dbl>         <dbl>
1 (Intercept)           83.7      3.54     23.7  0.00000000206
2 poly(AttPct, 3)1     -23.1     12.8      -1.81 0.104
3 poly(AttPct, 3)2     -24.7     12.8      -1.93 0.0851
4 poly(AttPct, 3)3     -44.3     12.8      -3.48 0.00699
```

```
glance(att_red_comp_pct_glm)$AIC
```

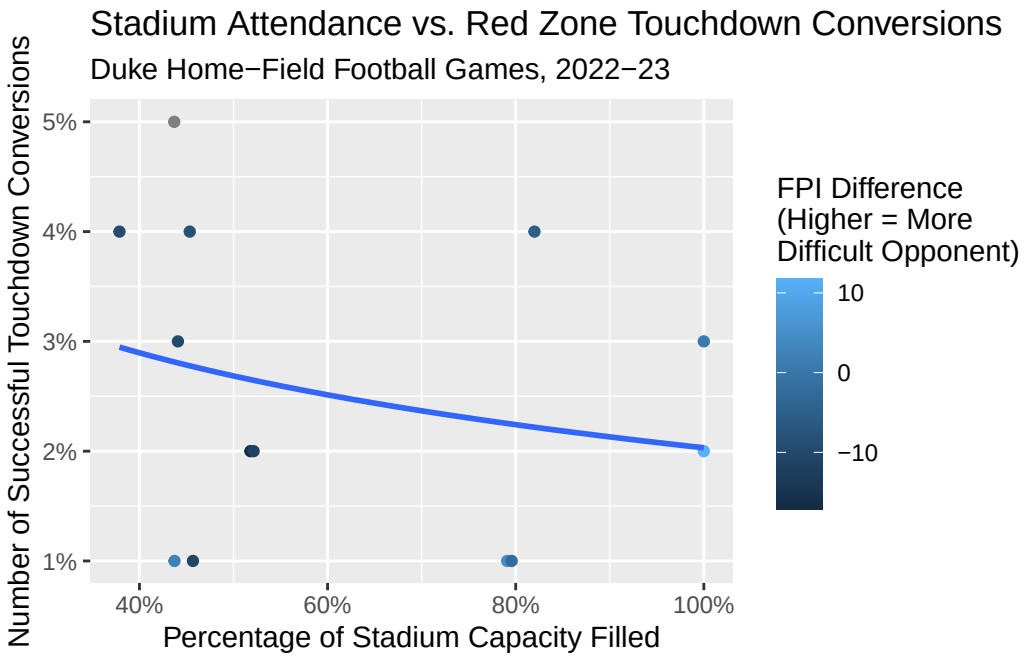
```
[1] 108.312
```

Based on Duke football games in Wallace Wade during 2022-23, there is *weak* evidence to suggest that red zone successful conversion percentage is associated with stadium attendance.

Attendance as a predictor of the number of successful red zone *touchdown* conversions:

```
# Visualization
home_off_red_conv_data |>
  ggplot(
    aes(x = AttPct, y = TD_Gained, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  scale_y_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Red Zone Touchdown Conversions",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Successful Touchdown Conversions",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_red_comp_touchdown_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(TD_Gained ~ log(AttPct), data = home_off_red_conv_data)

tidy(att_red_comp_touchdown_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    6.38       4.76     1.34    0.207
2 log(AttPct)   -0.944      1.17    -0.810   0.435
```

```
glance(att_red_comp_touchdown_glm)$AIC
```

```
[1] 49.68576
```

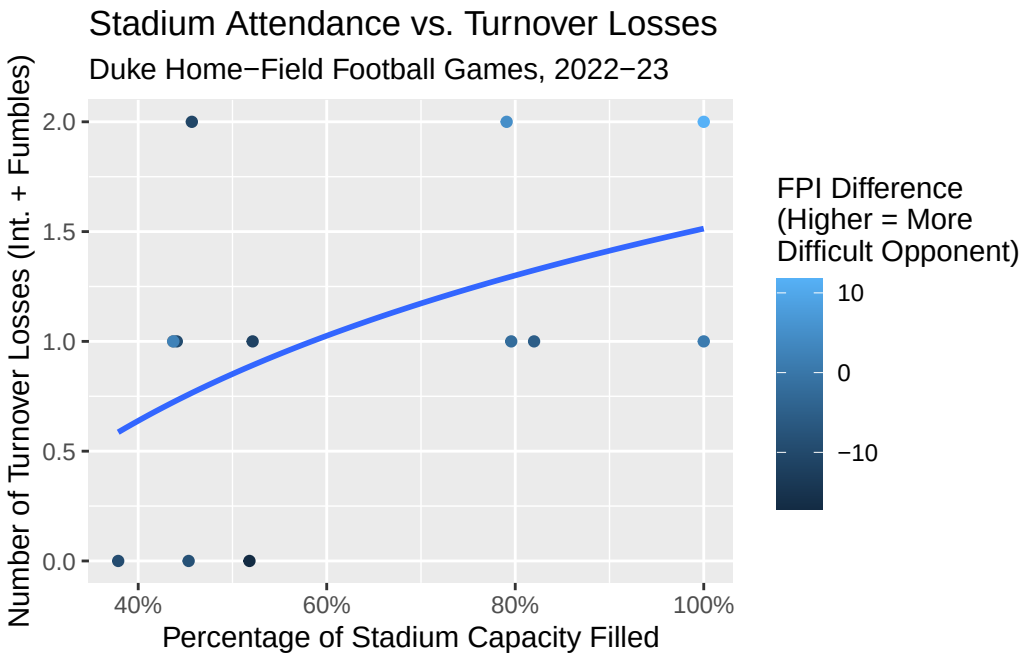
Wallace Wade attendance was *not* a statistically significant predictor of the number of successful red zone touchdown conversions in 2022-23.

Turnover Losses

Attendance as a predictor of the number of turnover losses by Duke per game:

```
# Dataset filtering
home_off_turnover_losses_data <- home_offense_data |>
  filter(Type == "Turnover_Losses")

# Visualization
home_off_turnover_losses_data |>
  ggplot(
    aes(x = AttPct, y = Comp, color = FPI_diff)
  ) +
  geom_point() +
  geom_smooth(method = "glm", formula = y ~ log(x), se = FALSE) +
  scale_x_continuous(labels = label_percent(scale = 1)) +
  labs(title = "Stadium Attendance vs. Turnover Losses",
       subtitle = "Duke Home-Field Football Games, 2022-23",
       x = "Percentage of Stadium Capacity Filled",
       y = "Number of Turnover Losses (Int. + Fumbles)",
       color = "FPI Difference\n(Higher = More\nDifficult Opponent)")
```

```
Warning: The following aesthetics were dropped during statistical transformation: colour
i This can happen when ggplot fails to infer the correct grouping structure in
  the data.
i Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```




```
# Linear model
att_turnover_losses_glm <- linear_reg() |>
  set_engine("glm") |>
  fit(Comp ~ log(AttPct), data = home_off_turnover_losses_data)

tidy(att_turnover_losses_glm)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   -2.89      2.19      -1.32   0.215
2 log(AttPct)    0.955     0.537      1.78   0.103
```

```
glance(att_turnover_losses_glm)$AIC
```

```
[1] 29.55889
```

Wallace Wade attendance was not a *strong* predictor of the number of Duke turnover losses per game in 2022-23.

Summary

On the whole, a relationship between stadium attendance at home and offensive performance did not appear to exist. The only truly significant finding was that, based on the data from 2022-23, when football attendance is 100% or very close to 100%, the predicted 3-point field goal success percentage decreases dramatically. This is likely due to factors other than the number of stadium attendees alone, such as the difficulty of the opponents at those games.