Predictions Based on 2011-2023 Home Attendance Records

This document seeks to utilize attendance records of Duke University home football games from the previous 12 seasons (2011-2023) to predict the number of attendees at Duke football home games during the 2024 season.

Packages

```
library(tidyverse)
library(tidymodels)
```

Importing the Dataset

```
att_data <- read_csv("data/Duke Stats - DukeAttendanceV3.csv")

att_data <- att_data |>
  mutate(Day = as.factor(Day))

home_att_data <- att_data |>
  filter(Site == "Home", Year < 2024)

home_opp_list <- c("Elon", "Connecticut", "Florida St.",
                   "North Carolina", "SMU", "Virginia Tech")

home_att_data
```

```
# A tibble: 84 x 51
   OppName       OppFPI DukeFPI FPI_diff DukeFPI_NetChange OppFPI_PrevYear
   <chr>          <dbl>   <dbl>    <dbl>              <dbl>           <dbl>
 1 Richmond        NA      -6.1     NA                -2.1            NA
 2 Stanford        24.4    -6.1     30.5              -2.1            24.2
 3 Tulane         -20.3    -6.1    -14.2              -2.1           -17.3
 4 Florida St.     15.3    -6.1     21.4              -2.1            17.2
 5 Wake Forest     -0.2    -6.1      5.9              -2.1            -6
 6 Virginia Tech   11.8    -6.1     17.9              -2.1            18.4
 7 Georgia Tech     5      -6.1     11.1              -2.1             5.3
 8 Florida Int'l   -8      -1.7     -6.3               4.4            -5.1
 9 N.C. Central    NA      -1.7     NA                 4.4            NA
10 Memphis        -13.2    -1.7    -11.5               4.4           -24.6
# i 74 more rows
# i 45 more variables: FPI_Diff_PrevYear <dbl>, Surface <chr>, Month <dbl>,
#   Date <dbl>, Year <dbl>, Day <fct>, Start_Time <dbl>, Site <chr>,
#   Result <chr>, DukePts <dbl>, OppPts <dbl>, PointDiff <dbl>, AttNum <dbl>,
#   AttPct <dbl>, ESPN_WinPred <dbl>, COVID_Limit <lgl>, Rain <lgl>,
#   City <chr>, State <chr>, TV_Coverage <chr>, Bowl <lgl>,
#   DukeRankGametime <dbl>, OppRankGametime <dbl>, OppRankSeasonEnd <dbl>, ...
```

Attendance History for 2024 Opponents

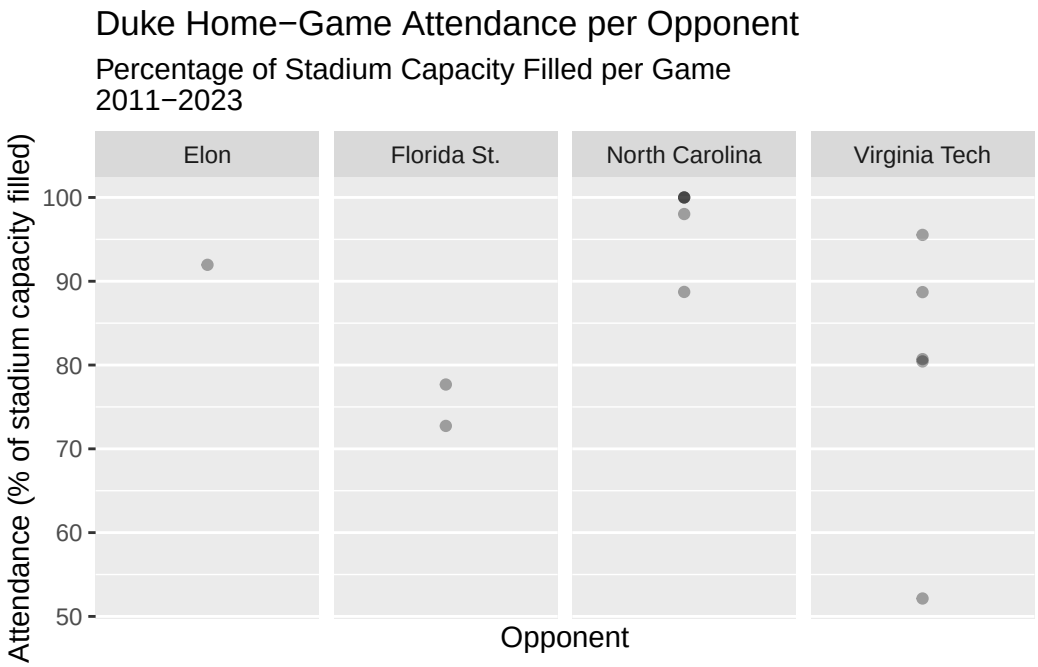
Wallace Wade Stadium capacity:

- *Pre-rennovation:* 33,941 (1982–2015)
- *Post-rennovation:* 40,004 (2016-present)

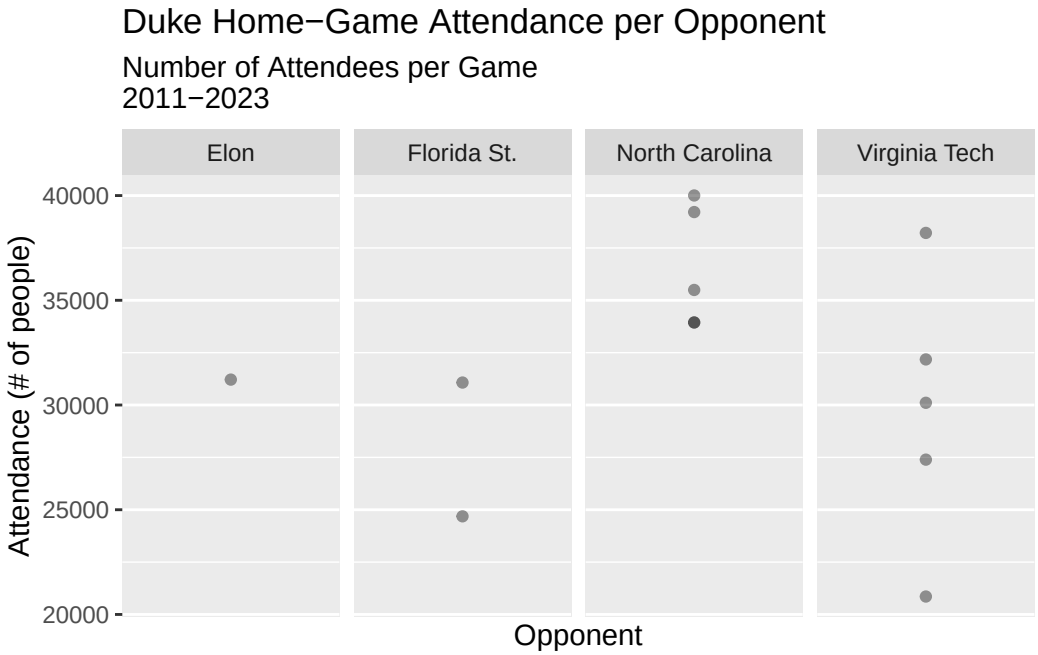
All Teams

```
home_att_data |>
  filter(OppName %in% home_opp_list) |>
  ggplot(
    aes(x = 0, y = AttPct)
  ) +
  geom_point(alpha = 0.333) +
  facet_wrap(~OppName, strip.position = "top", nrow = 1) +
  scale_x_continuous(labels = NULL, breaks = NULL) +
  labs(title = "Duke Home-Game Attendance per Opponent",
```

```
      subtitle = "Percentage of Stadium Capacity Filled per Game\n2011-2023",
      x = "Opponent",
      y = "Attendance (% of stadium capacity filled)")
```




```
home_att_data |>
  filter(OppName %in% home_opp_list) |>
  ggplot(
    aes(x = 0, y = AttNum)
  ) +
  geom_point(alpha = 0.4) +
  facet_wrap(~OppName, strip.position = "top", nrow = 1) +
  scale_x_continuous(labels = NULL, breaks = NULL) +
  labs(title = "Duke Home-Game Attendance per Opponent",
       subtitle = "Number of Attendees per Game\n2011-2023",
       x = "Opponent",
       y = "Attendance (# of people)")
```



Elon

```
home_att_data |>
  filter(OppName == "Elon") |>
  summarize("Name" = OppName,
            "End-of-Season FPI" = OppFPI,
            Month,
            Date,
            Year,
            "# of Attendees" = AttNum,
            "% of Stadium Capacity Filled" = AttPct)
```

```
# A tibble: 1 x 7
  Name  `End-of-Season FPI` Month  Date  Year `# of Attendees`
  <chr>               <dbl> <dbl> <dbl> <dbl>            <dbl>
1 Elon                   NA     8    30  2014            31213
# i 1 more variable: `% of Stadium Capacity Filled` <dbl>
```

Connecticut

```
home_att_data |>
  filter(OppName == "Connecticut") |>
  summarize("Name" = OppName,
            "End-of-Season FPI" = OppFPI,
            Month,
            Date,
            Year,
            "# of Attendees" = AttNum,
            "% of Stadium Capacity Filled" = AttPct)
```

```
# A tibble: 0 x 7
# i 7 variables: Name <chr>, End-of-Season FPI <dbl>, Month <dbl>, Date <dbl>,
#   Year <dbl>, # of Attendees <dbl>, % of Stadium Capacity Filled <dbl>
```

UConn never faced against Duke in Wallace Wade Stadium from 2011 to 2023.

Florida St.

```
home_att_data |>
  filter(OppName == "Florida St.") |>
  summarize("Name" = OppName,
            "End-of-Season FPI" = OppFPI,
            Month,
            Date,
            Year,
            "# of Attendees" = AttNum,
            "% of Stadium Capacity Filled" = AttPct)
```

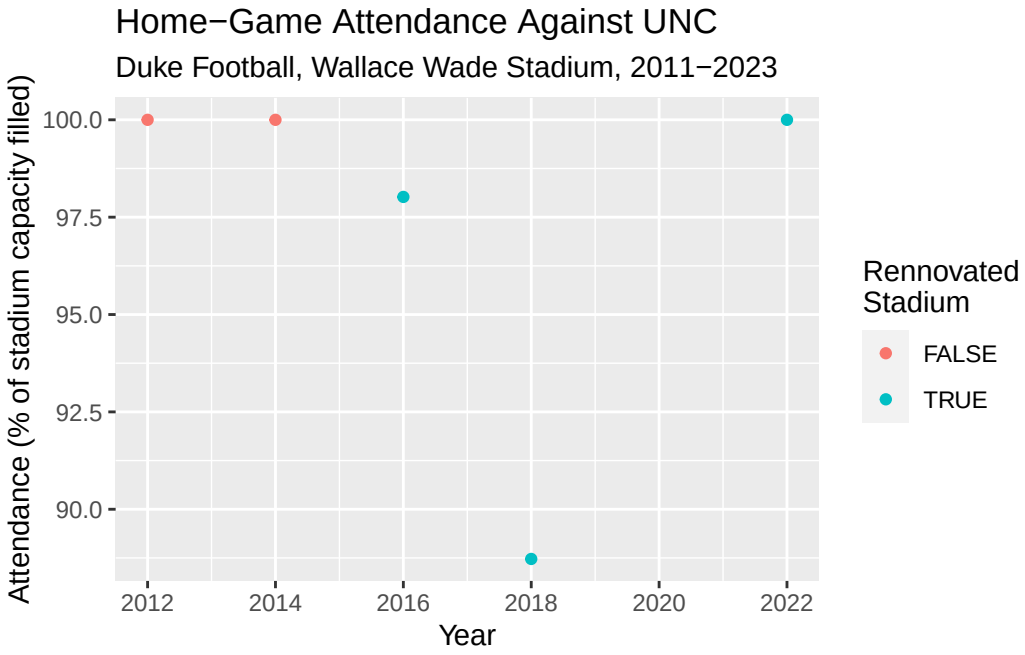
```
# A tibble: 2 x 7
  Name        `End-of-Season FPI` Month  Date  Year `# of Attendees`
  <chr>                     <dbl> <dbl> <dbl> <dbl>            <dbl>
1 Florida St.                15.3    10    15  2011            24687
2 Florida St.                13.3    10    14  2017            31073
# i 1 more variable: `% of Stadium Capacity Filled` <dbl>
```

North Carolina

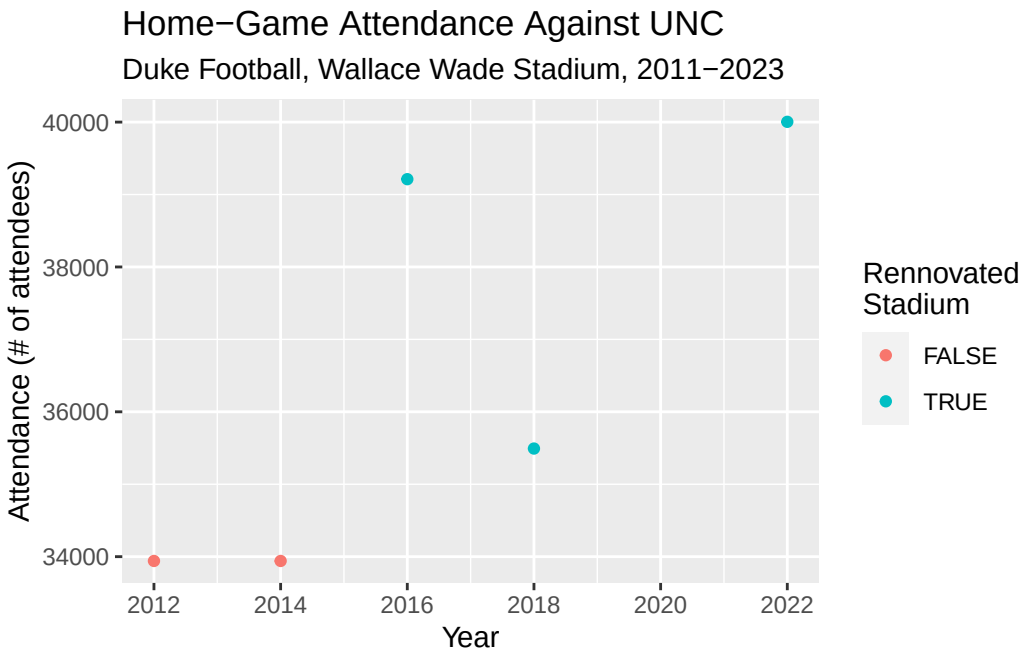
```
home_att_data |>
  filter(OppName == "North Carolina") |>
  summarize("Name" = OppName,
            "End-of-Season FPI" = OppFPI,
            Month,
            Date,
            Year,
            "# of Attendees" = AttNum,
            "% of Stadium Capacity Filled" = AttPct)
```

```
# A tibble: 6 x 7
  Name           `End-of-Season FPI` Month  Date  Year `# of Attendees`
  <chr>                        <dbl> <dbl> <dbl> <dbl>            <dbl>
1 North Carolina                10.6    10    20  2012            33941
2 North Carolina                 4.4    11    20  2014            33941
3 North Carolina                14      11    10  2016            39212
4 North Carolina                -2.6    11    10  2018            35493
5 North Carolina                10.2    11     7  2020               NA
6 North Carolina                 6.2    10    15  2022            40004
# i 1 more variable: `% of Stadium Capacity Filled` <dbl>
```

```
home_att_data |>
  filter(OppName == "North Carolina") |>
  ggplot(
    aes(x = Year, y = AttPct, color = Rennovated)
  ) +
  geom_point() +
  scale_x_continuous(breaks = seq(from = 2012, to = 2023, by = 2)) +
  labs(title = "Home-Game Attendance Against UNC",
       subtitle = "Duke Football, Wallace Wade Stadium, 2011-2023",
       x = "Year",
       y = "Attendance (% of stadium capacity filled)",
       color = "Rennovated\nStadium")
```




```
home_att_data |>
  filter(OppName == "North Carolina") |>
  ggplot(
    aes(x = Year, y = AttNum, color = Rennovated)
  ) +
  geom_point() +
  scale_x_continuous(breaks = seq(from = 2012, to = 2023, by = 2)) +
  labs(title = "Home-Game Attendance Against UNC",
       subtitle = "Duke Football, Wallace Wade Stadium, 2011-2023",
       x = "Year",
       y = "Attendance (# of attendees)",
       color = "Rennovated\nStadium")
```



SMU

```
home_att_data |>
  filter(OppName == "SMU") |>
  summarize("Name" = OppName,
            "End-of-Season FPI" = OppFPI,
            Month,
            Date,
            Year,
            "# of Attendees" = AttNum,
            "% of Stadium Capacity Filled" = AttPct)
```

```
# A tibble: 0 x 7
# i 7 variables: Name <chr>, End-of-Season FPI <dbl>, Month <dbl>, Date <dbl>,
#   Year <dbl>, # of Attendees <dbl>, % of Stadium Capacity Filled <dbl>
```

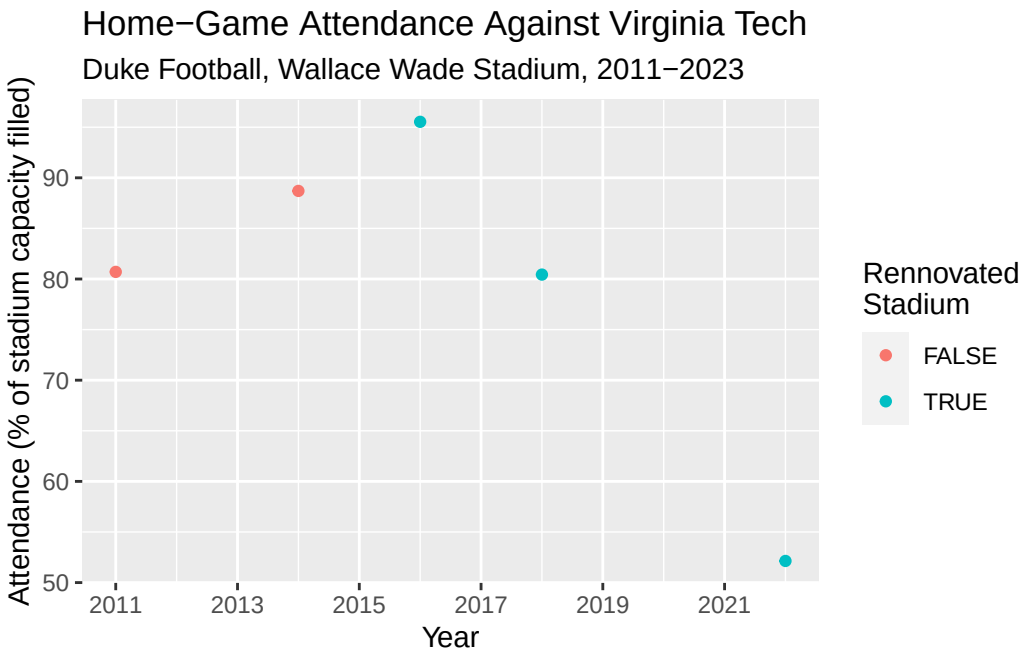
UConn never faced against Duke in Wallace Wade Stadium from 2011 to 2023.

Virginia Tech

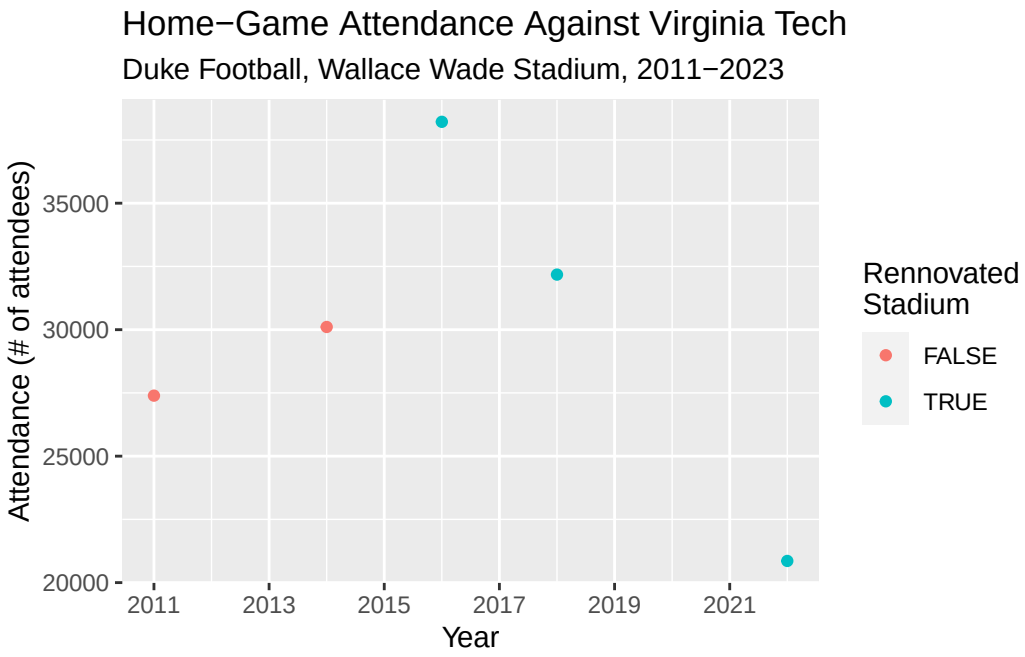
```
home_att_data |>
  filter(OppName == "Virginia Tech") |>
  summarize("Name" = OppName,
            "End-of-Season FPI" = OppFPI,
            Month,
            Date,
            Year,
            "# of Attendees" = AttNum,
            "% of Stadium Capacity" = AttPct)
```

```
# A tibble: 6 x 7
  Name          `End-of-Season FPI` Month  Date  Year `# of Attendees`
  <chr>                       <dbl> <dbl> <dbl> <dbl>            <dbl>
1 Virginia Tech                11.8    10    29  2011            27392
2 Virginia Tech                 7.9    11    15  2014            30107
3 Virginia Tech                13.7    11     5  2016            38217
4 Virginia Tech                 3.4     9    29  2018            32177
5 Virginia Tech                 7.3    10     3  2020               NA
6 Virginia Tech                -6.2    11    12  2022            20857
# i 1 more variable: `% of Stadium Capacity` <dbl>
```

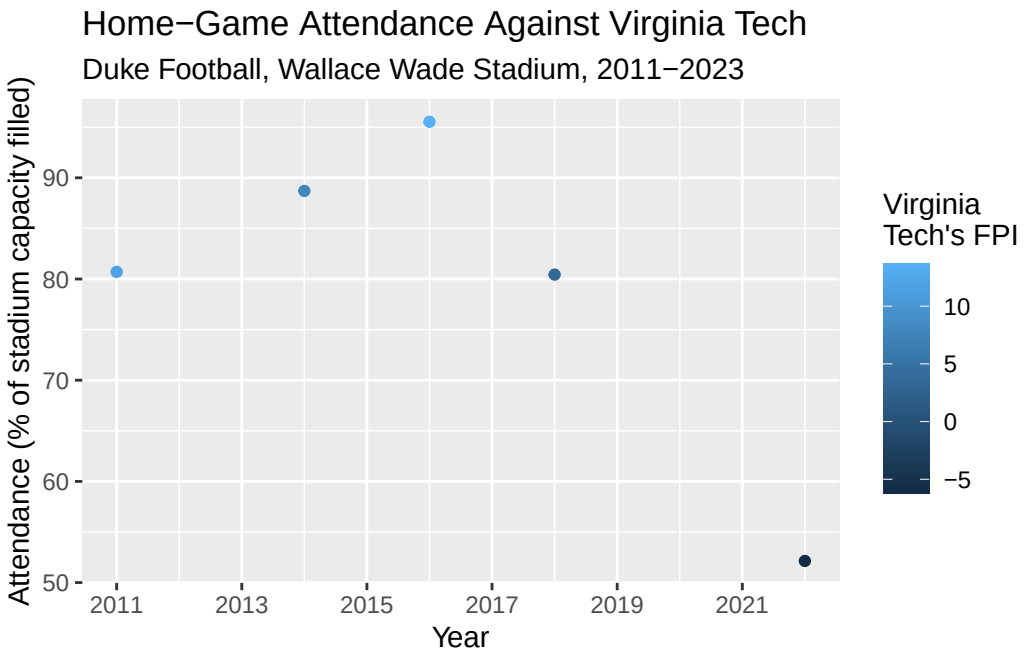
```
home_att_data |>
  filter(OppName == "Virginia Tech") |>
  ggplot(
    aes(x = Year, y = AttPct, color = Rennovated)
  ) +
  geom_point() +
  scale_x_continuous(breaks = seq(from = 2011, to = 2023, by = 2)) +
  labs(title = "Home-Game Attendance Against Virginia Tech",
       subtitle = "Duke Football, Wallace Wade Stadium, 2011-2023",
       x = "Year",
       y = "Attendance (% of stadium capacity filled)",
       color = "Rennovated\nStadium")
```




```
home_att_data |>
  filter(OppName == "Virginia Tech") |>
  ggplot(
    aes(x = Year, y = AttNum, color = Rennovated)
  ) +
  geom_point() +
  scale_x_continuous(breaks = seq(from = 2011, to = 2023, by = 2)) +
  labs(title = "Home-Game Attendance Against Virginia Tech",
       subtitle = "Duke Football, Wallace Wade Stadium, 2011-2023",
       x = "Year",
       y = "Attendance (# of attendees)",
       color = "Rennovated\nStadium")
```




```
home_att_data |>
  filter(OppName == "Virginia Tech") |>
  ggplot(
    aes(x = Year, y = AttPct, color = OppFPI)
  ) +
  geom_point() +
  scale_x_continuous(breaks = seq(from = 2011, to = 2023, by = 2)) +
  labs(title = "Home-Game Attendance Against Virginia Tech",
       subtitle = "Duke Football, Wallace Wade Stadium, 2011-2023",
       x = "Year",
       y = "Attendance (% of stadium capacity filled)",
       color = "Virginia\nTech's FPI")
```


Home−Game Attendance Against Virginia Tech
Duke Football, Wallace Wade Stadium, 2011−2023
Attendance (% of stadium capacity filled)
90
80
70
60
50
2011
2013
2015
2017
2019
2021
Year
Virginia Tech's FPI
10
5
0
−5