

# 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)
```

```
-- 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 become
```

```
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()
* Dig deeper into tidy modeling with R at https://www.tmwr.org

```

## Importing the Dataset

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

Rows: 176 Columns: 51

```

-- Column specification -----
Delimiter: ","
chr  (8): OppName, Surface, Day, Site, Result, City, State, TV_Coverage
dbl (24): OppFPI, DukeFPI, FPI_diff, DukeFPI_NetChange, OppFPI_PrevYear, FPI...
lgl (19): COVID_Limit, Rain, Bowl, DukeRankedGametime, OppRankedGametime, Op...

```

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.

```

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

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

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

home_att_data

```

# A tibble: 86 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 76 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>, ...
```

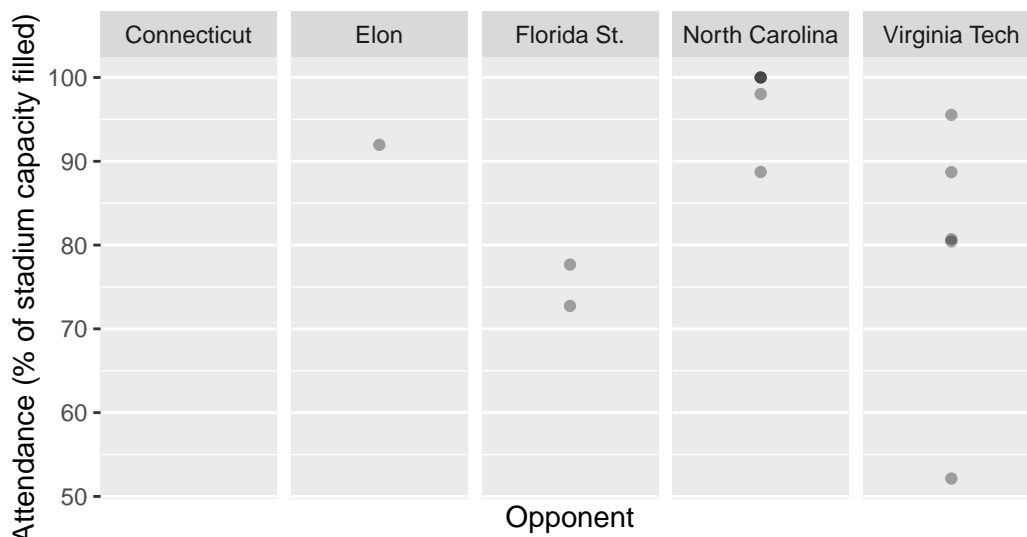
## Attendance History for 2024 Opponents

```
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)")
```

Warning: Removed 4 rows containing missing values (`geom\_point()`).

## Duke Home-Game Attendance per Opponent

Percentage of Stadium Capacity Filled per Game  
2011-2023

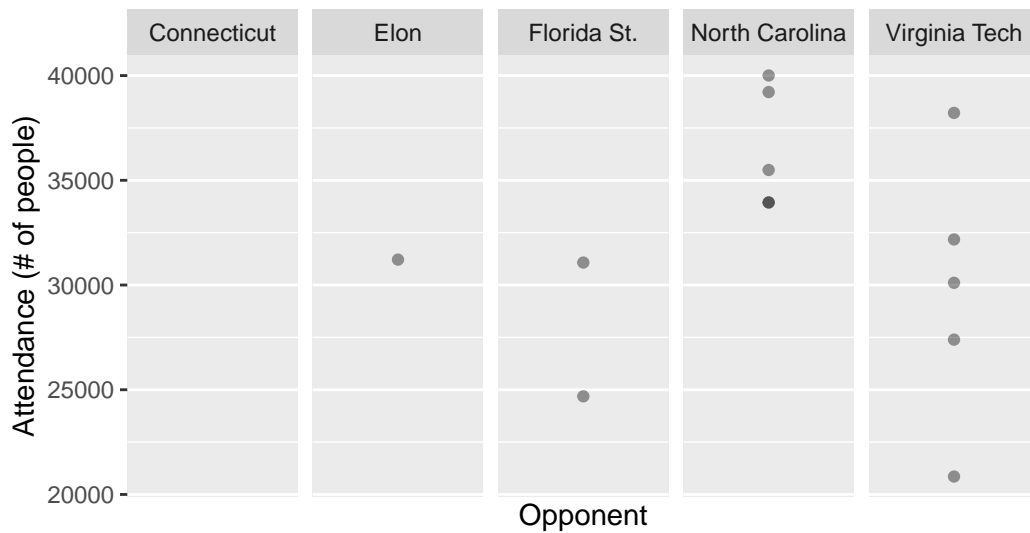


```
home_att_data |>
  filter(OppName %in% home_opp_list) |>
  ggplot(
    aes(x = 0, y = AttNum)
  ) +
  geom_point(alpha = 0.4) +
  facet_wrap(~OppName, strip.position = "top", nrow = 1) +
  scale_x_continuous(labels = NULL, breaks = NULL) +
  labs(title = "Duke Home-Game Attendance per Opponent",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

Warning: Removed 4 rows containing missing values (`geom\_point()`).

## Duke Home–Game Attendance per Opponent

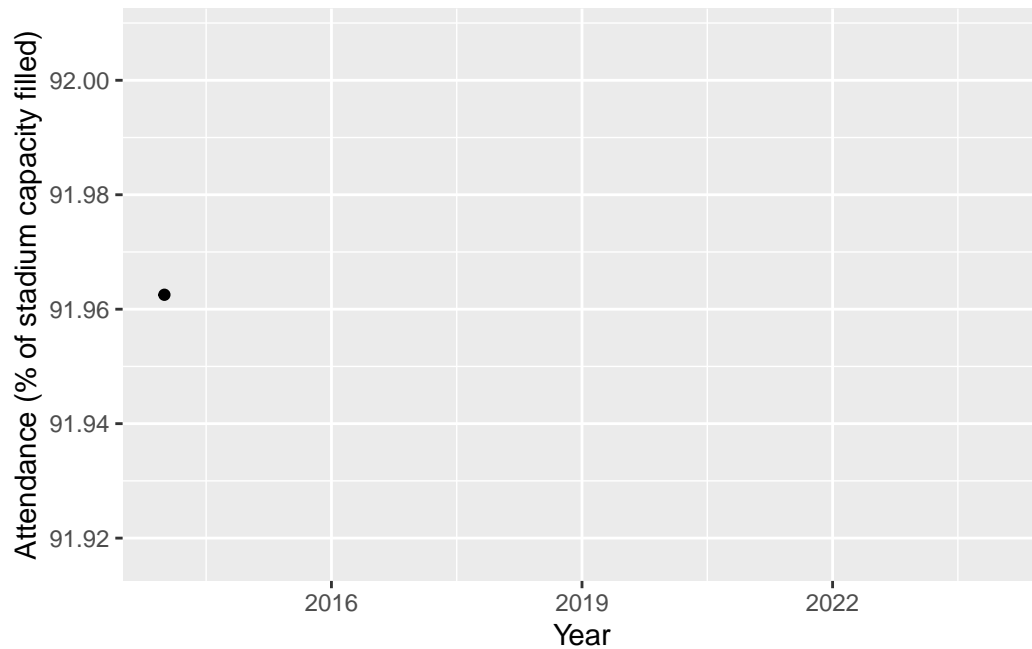
Number of Attendees per Game  
2011–2023



### Elon

```
home_att_data |>
  filter(OppName == "Elon") |>
  ggplot(
    aes(x = Year, y = AttPct)
  ) +
  geom_point() +
  labs(x = "Year",
       y = "Attendance (% of stadium capacity filled)")
```

Warning: Removed 1 rows containing missing values (`geom\_point()`).



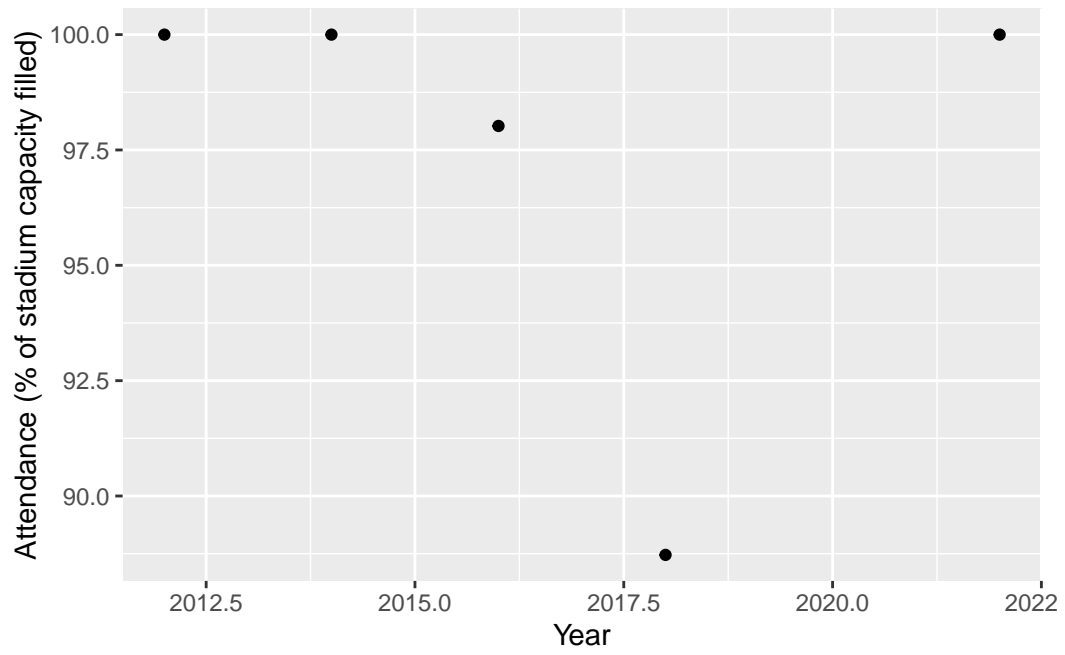
**Connecticut**

**Florida St.**

**North Carolina**

```
home_att_data |>
  filter(OppName == "North Carolina") |>
  ggplot(
    aes(x = Year, y = AttPct)
  ) +
  geom_point() +
  labs(x = "Year",
       y = "Attendance (% of stadium capacity filled)")
```

Warning: Removed 1 rows containing missing values (`geom\_point()`).



**SMU**

**Virginia Tech**