Data Science Projects in the Florida Panthers Organization

Brian Macdonald

Florida Panthers Hockey Club, Sunrise, FL

LADS 2018, New College of Florida

Twitter: @bmacNHL, @FlaPanthers LinkedIn: bmacNHL

Outline

- One specific example that involves hockey and business data

Outline

- One specific example that involves hockey and business data
- Overview of other ways we use data
  - Renewal prediction
  - Arena maps
  - South Florida maps

Outline

- One specific example that involves hockey and business data
- Overview of other ways we use data
  - Renewal prediction
  - Arena maps
  - South Florida maps
- Future areas of focus

Attendance Model

Goal:

- Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

Attendance Model

Goal:

- Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

This could help answer questions like:

- Which games should be in which tiers for variable pricing?

Attendance Model

Goal:

- Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

This could help answer questions like:

- Which games should be in which tiers for variable pricing?

- What kinds of things could we request when the league is developing the schedule?

Attendance Model

Goal:

- Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

This could help answer questions like:

- Which games should be in which tiers for variable pricing?
- What kinds of things could we request when the league is developing the schedule?
    - **Specific question**: Do we prefer good team on a Saturday and bad team during the week, or a good team during the week and a bad team on Saturday?"

Attendance Model

Goal:

- Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

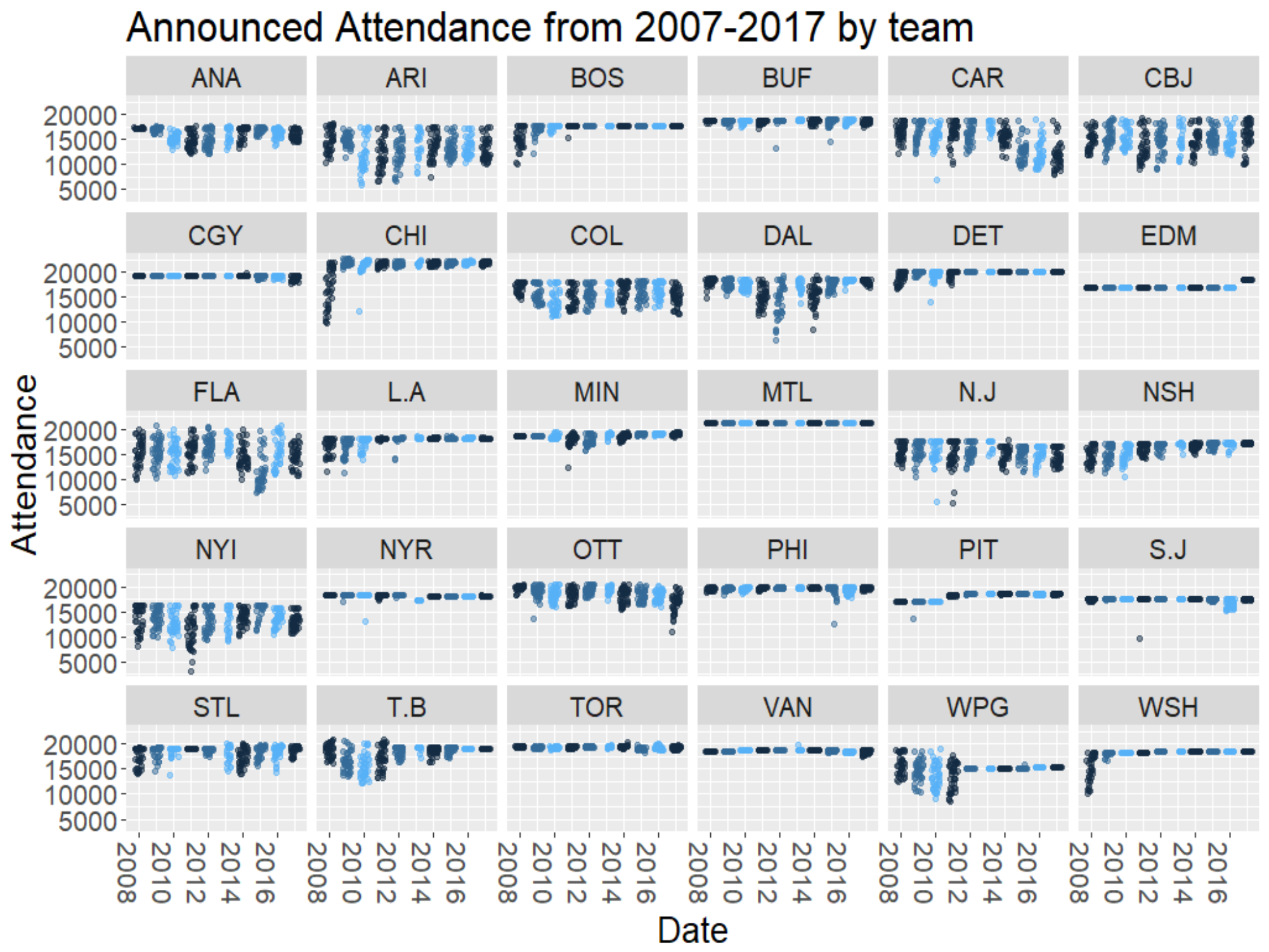
This could help answer questions like:

- Which games should be in which tiers for variable pricing?
- What kinds of things could we request when the league is developing the schedule?
  - **Specific question**: Do we prefer good team on a Saturday and bad team during the week, or a good team during the week and a bad team on Saturday?"
  - What do we want Thanksgiving week?

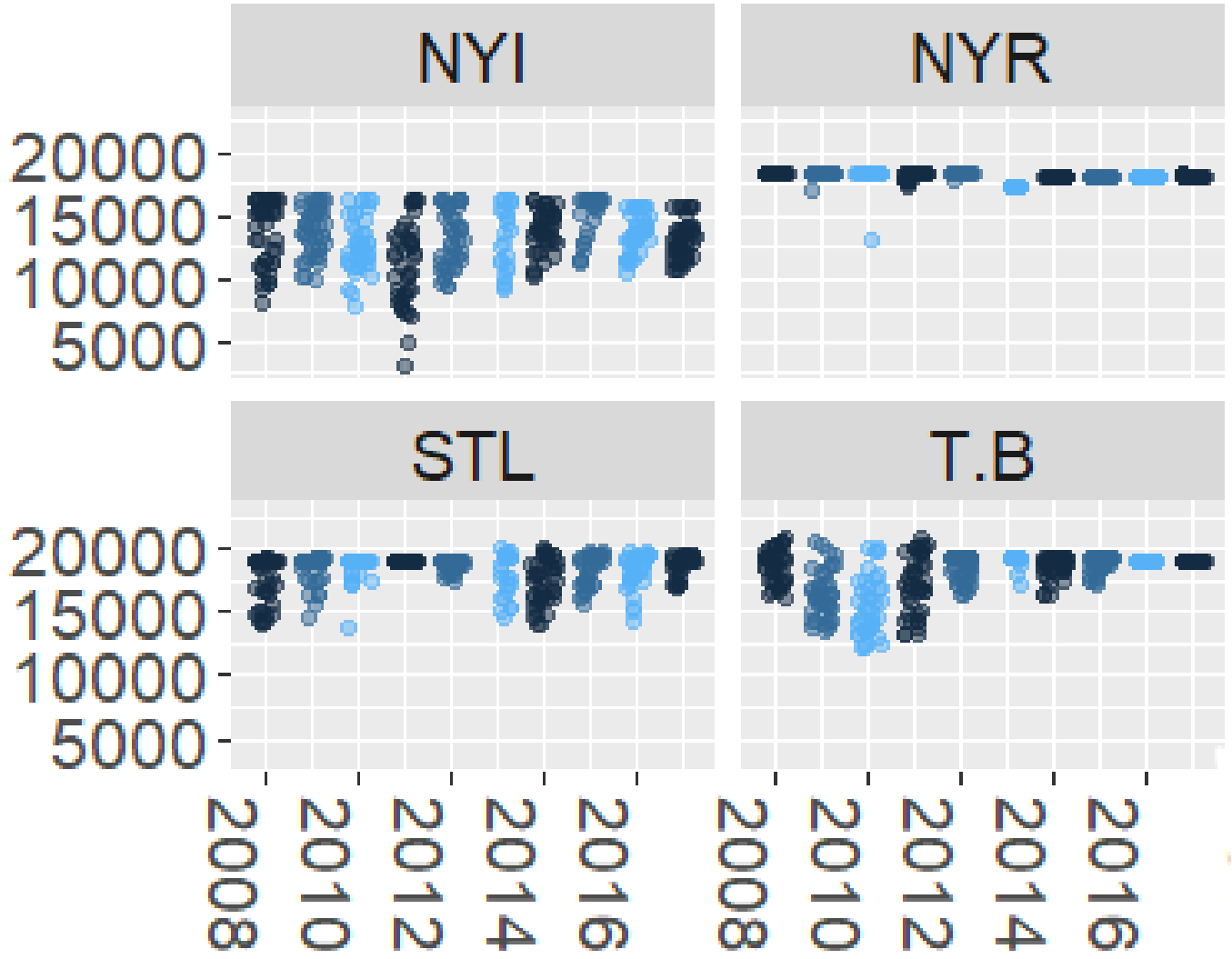
First, let’s plot some raw data.
Attendance* by game, from 2007-08 to 2016-17, for all 30 teams.

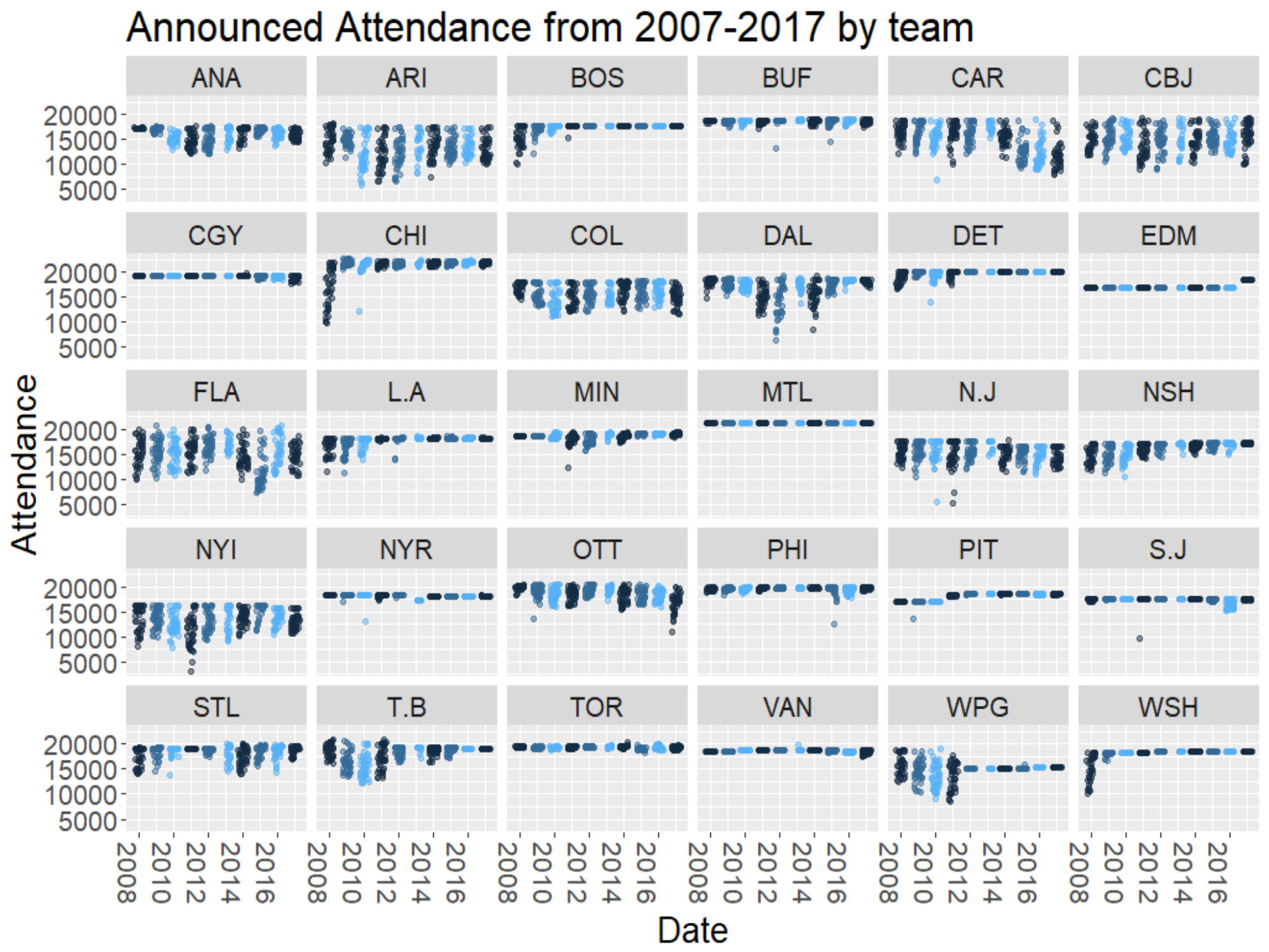
First, let's plot some raw data.
Attendance* by game, from 2007-08 to 2016-17, for all 30 teams.

*Announced attendance, as published on nhl.com


Announced Attendance from 2007-2017 by team
ANA
ARI
BOS
BUF
CAR
CBJ
CGY
CHI
COL
DAL
DET
EDM
FLA
L.A
MIN
MTL
N.J
NSH
NYI
NYR
OTT
PHI
PIT
S.J
STL
T.B
TOR
VAN
WPG
WSH
Attendance
20000
15000
10000
5000
2008
2010
2012
2014
2016
Date

Snippet


NYI
NYR
STL
T.B
20000
15000
10000
5000
2008
2010
2012
2014
2016


Announced Attendance from 2007-2017 by team
ANA
ARI
BOS
BUF
CAR
CBJ
CGY
CHI
COL
DAL
DET
EDM
FLA
L.A
MIN
MTL
N.J
NSH
NYI
NYR
OTT
PHI
PIT
S.J
STL
T.B
TOR
VAN
WPG
WSH
Attendance
20000
15000
10000
5000
2008
2010
2012
2014
2016
Date

Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.

Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.

Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.
3. Remove a few games unusual characteristics.
   - European games

Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.
3. Remove a few games unusual characteristics.
   - European games
   - Blizzards

Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.
3. Remove a few games unusual characteristics.
   - European games
   - Blizzards
4. Use several predictor variables (next slide)

Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.
3. Remove a few games unusual characteristics.
   - European games
   - Blizzards
4. Use several predictor variables (next slide)
5. Announced attendance is outcome we're trying to predict

Predictors

- home team, away team

Predictors

- home team, away team
- day of week, month

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)
- season opener (Y or N)

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)
- season opener (Y or N)
- same division (Y or N)

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)
- season opener (Y or N)
- same division (Y or N)
- same conference (Y or N)

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)
- season opener (Y or N)
- same division (Y or N)
- same conference (Y or N)
- points during previous year for home/away (lag)

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)
- season opener (Y or N)
- same division (Y or N)
- same conference (Y or N)
- points during previous year for home/away (lag)
- year-to-date points relative to average for home/away.

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)
- season opener (Y or N)
- same division (Y or N)
- same conference (Y or N)
- points during previous year for home/away (lag)
- year-to-date points relative to average for home/away.
- day and month interaction (Sundays different in fall?)

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)
- season opener (Y or N)
- same division (Y or N)
- same conference (Y or N)
- points during previous year for home/away (lag)
- year-to-date points relative to average for home/away.
- day and month interaction (Sundays different in fall?)
- home team and day interaction

Predictors

- home team, away team
- day of week, month
- holiday (Columbus Day, Thanksgiving week, etc., or none.)
- season opener (Y or N)
- same division (Y or N)
- same conference (Y or N)
- points during previous year for home/away (lag)
- year-to-date points relative to average for home/away.
- day and month interaction (Sundays different in fall?)
- home team and day interaction
- home team and month interaction
  (snowbird months good for us?)

Interpretation of regression model results

- Impact that each of these variables have on attendance,

Interpretation of regression model results

- Impact that each of these variables have on attendance, **independent of all other variables**.

Interpretation of regression model results

- Impact that each of these variables have on attendance, **independent of all other variables**.

- For example, we find the effect of day, controlling for all of the other variables in our model

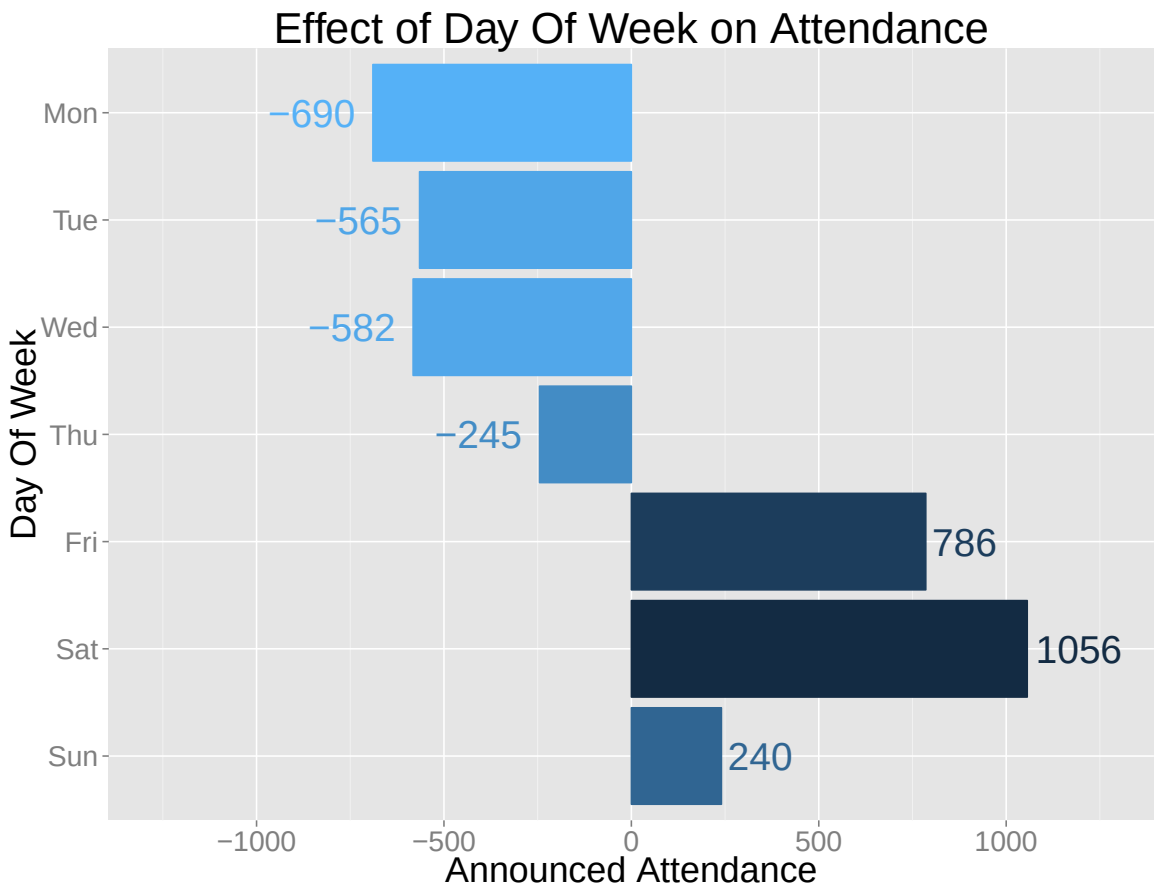
Interpretation of regression model results

- Impact that each of these variables have on attendance, **independent of all other variables**.

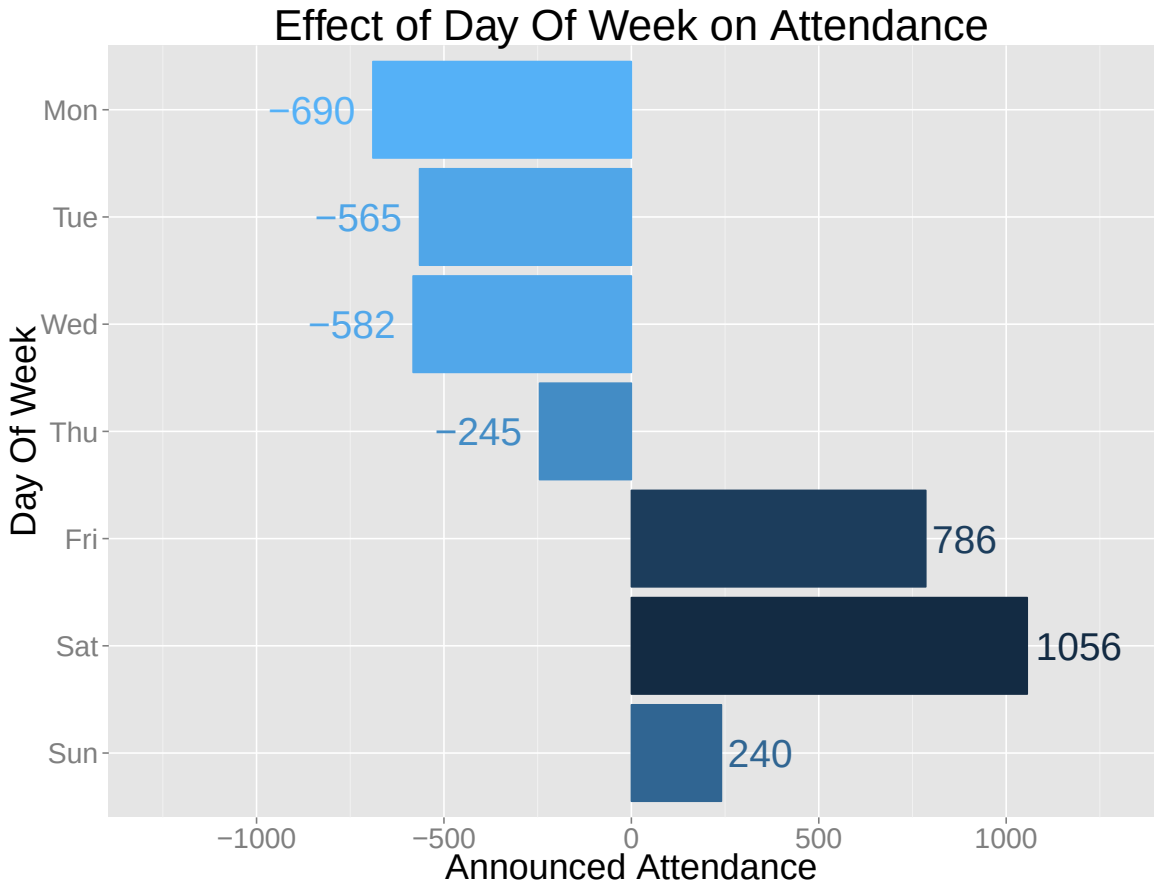
- For example, we find the effect of day, controlling for all of the other variables in our model

- That's an important point. Example: If teams schedule big opponents on the weekend, then the effect of a weekend game could be overstated if we just look at day and ignore opponent.

Example: day of week

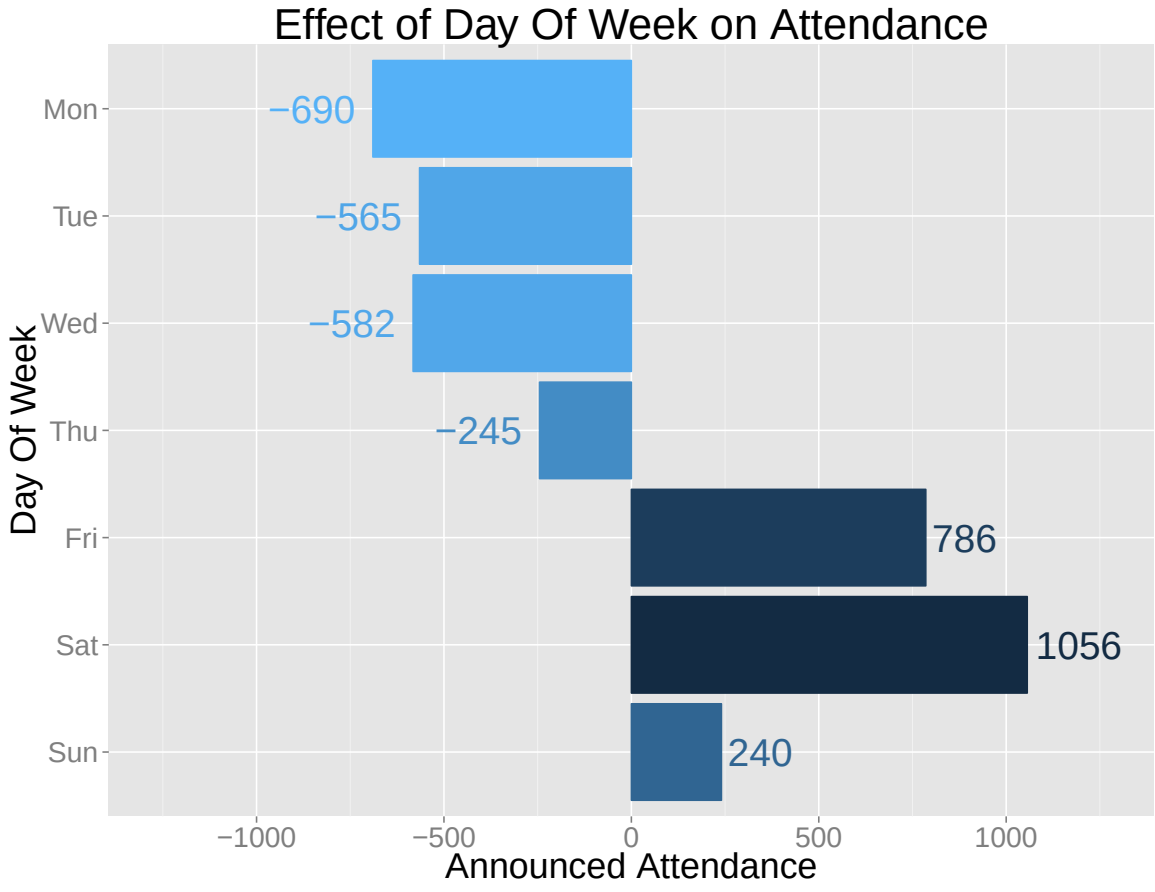


Example: day of week



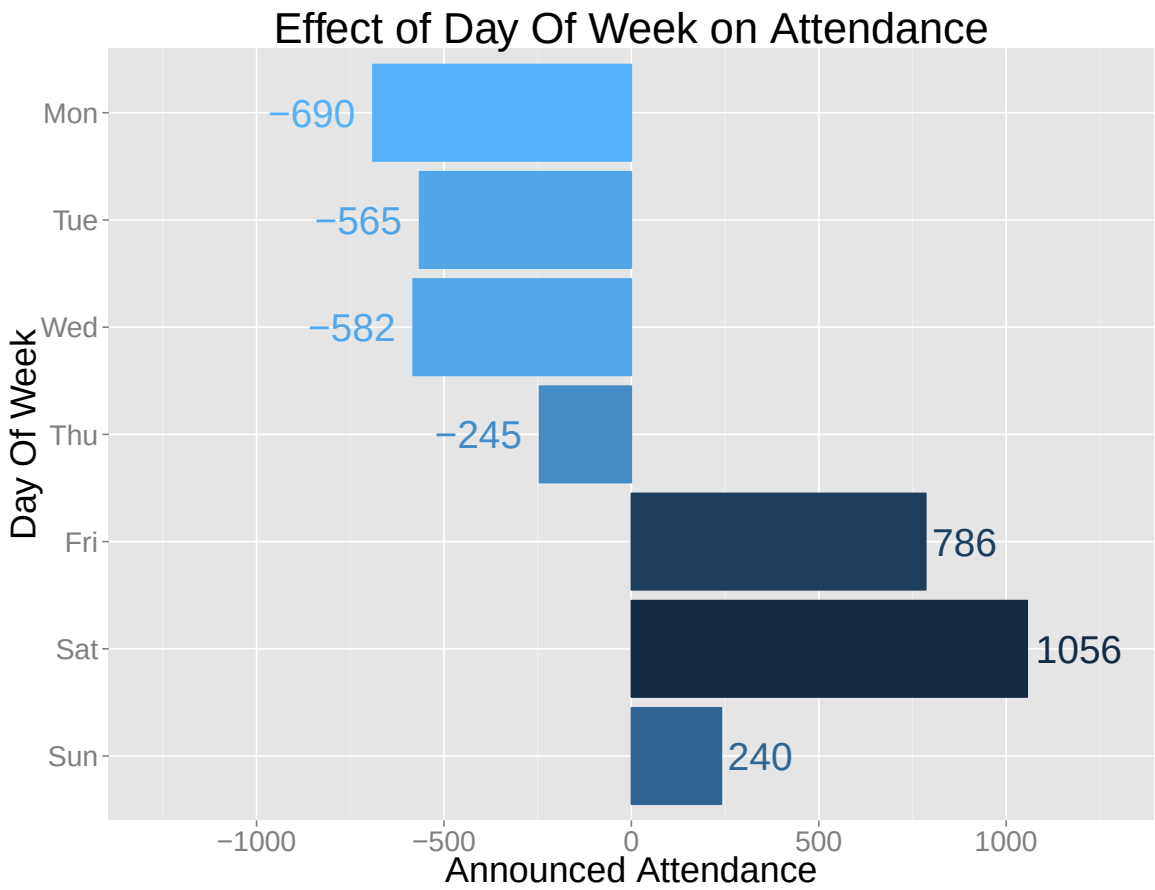
1. Attendance on Saturday is expected to be 1,056 higher than average, “holding all other variables constant."

Example: day of week



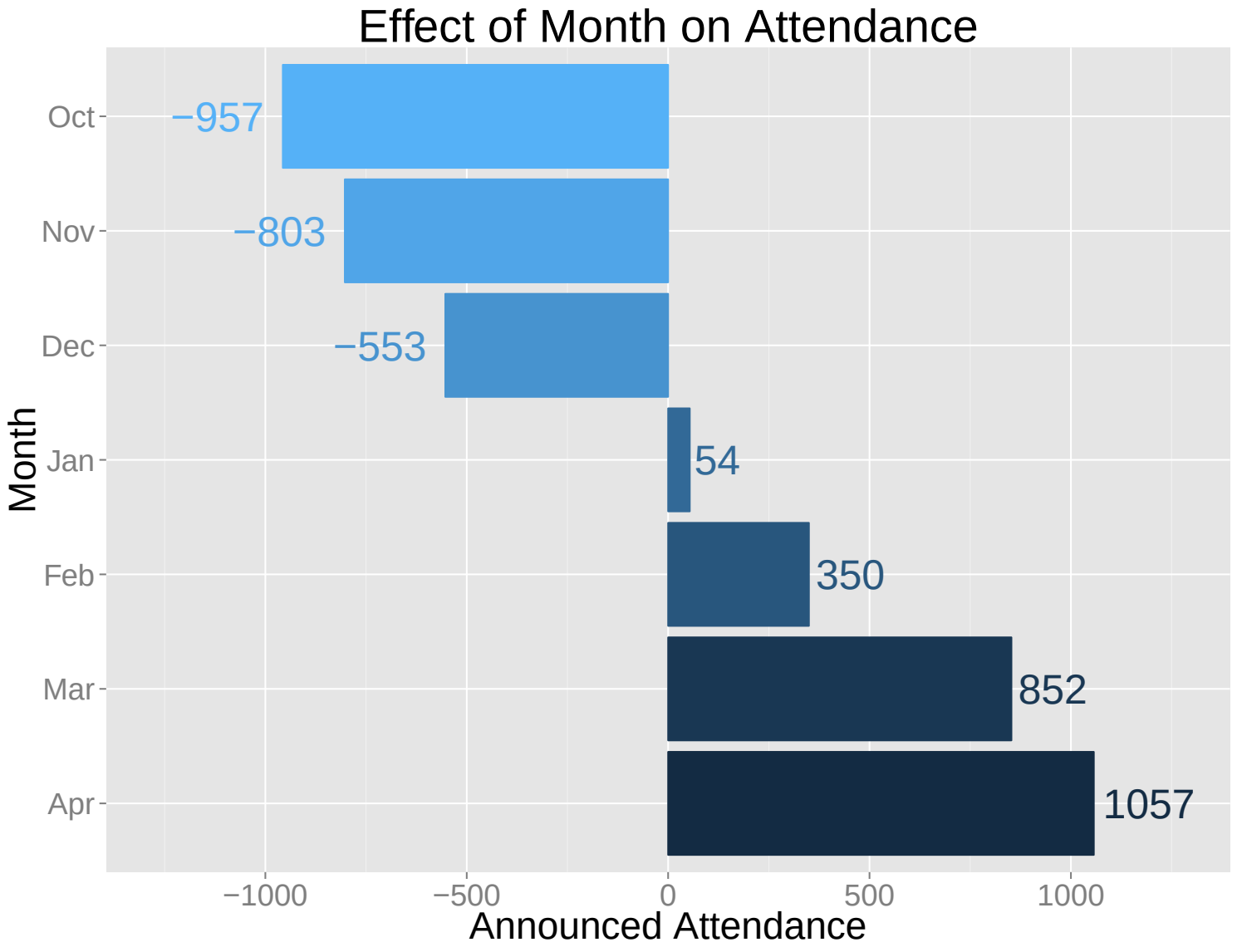
1. Attendance on Saturday is expected to be 1,056 higher than average, “holding all other variables constant."
2. The difference between Saturday and Monday is expected to be 1,746 (1,056 + 690).

Example: day of week

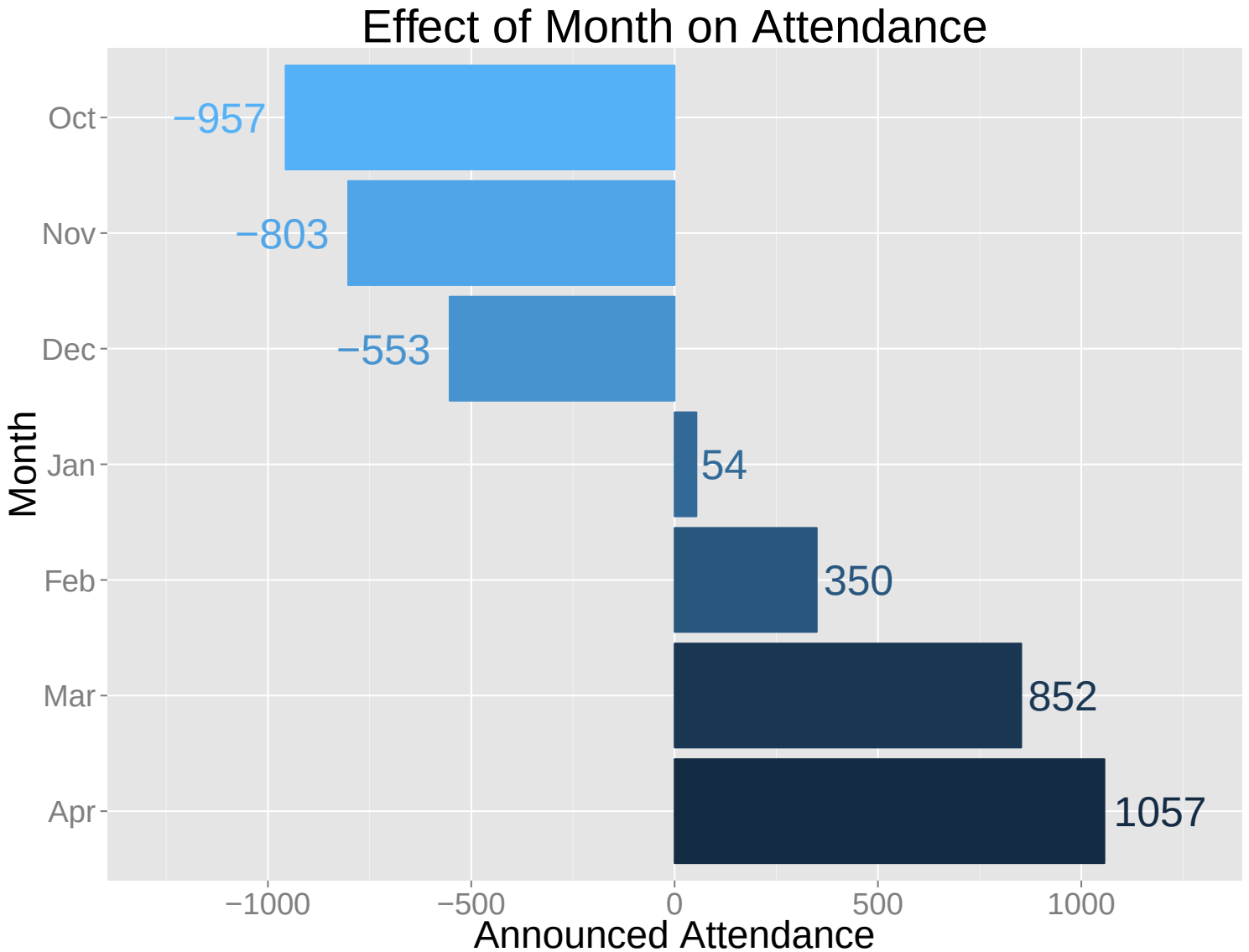


1. Attendance on Saturday is expected to be 1,056 higher than average, “holding all other variables constant."
2. The difference between Saturday and Monday is expected to be 1,746 (1,056 + 690).
3. Not surprising. Stuff we knew. But now we’ve quantified.

Month

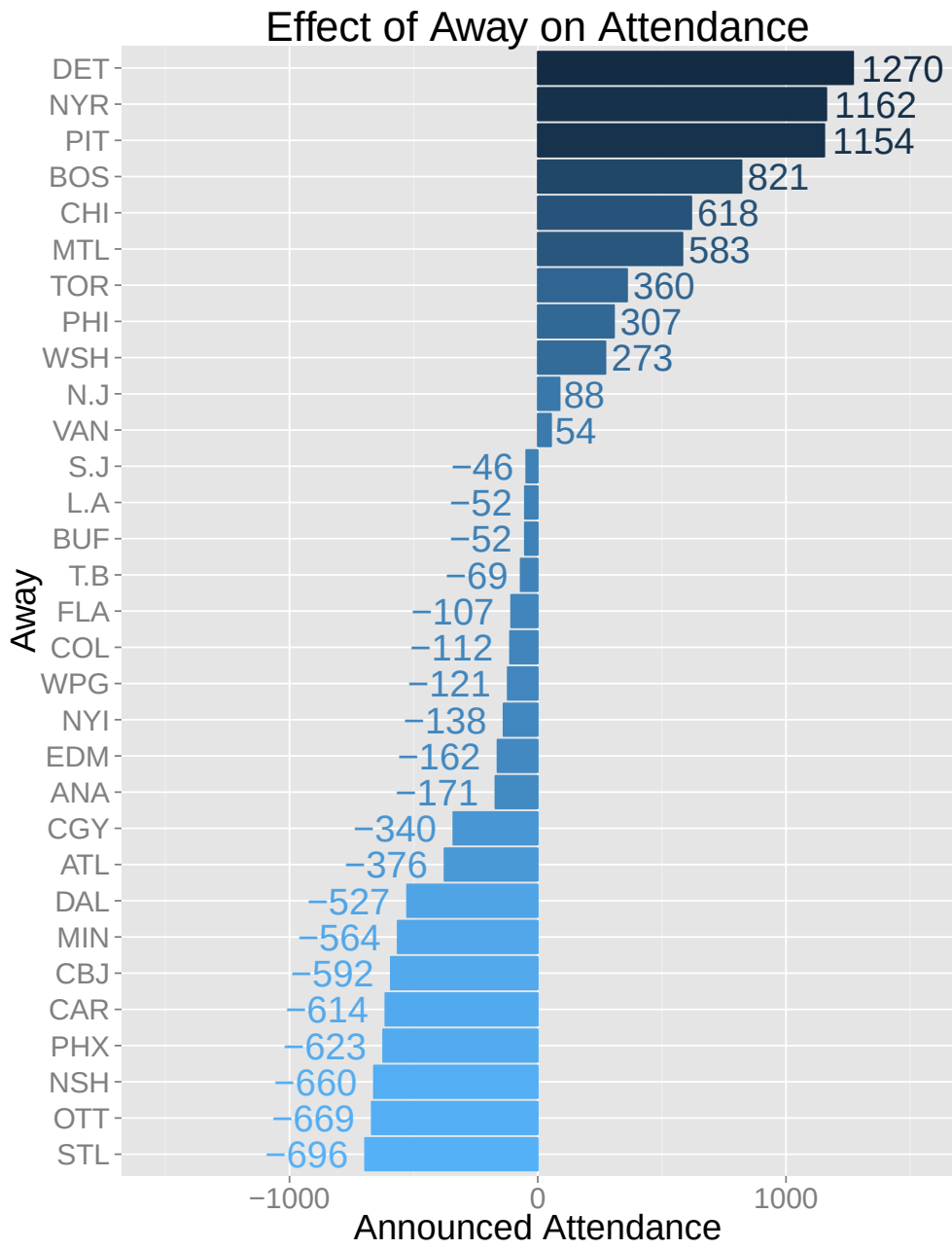

Effect of Month on Attendance
Oct
Nov
Dec
Jan
Feb
Mar
Apr
Month
−957
−803
−553
54
350
852
1057
−1000
−500
0
500
1000
Announced Attendance

Month

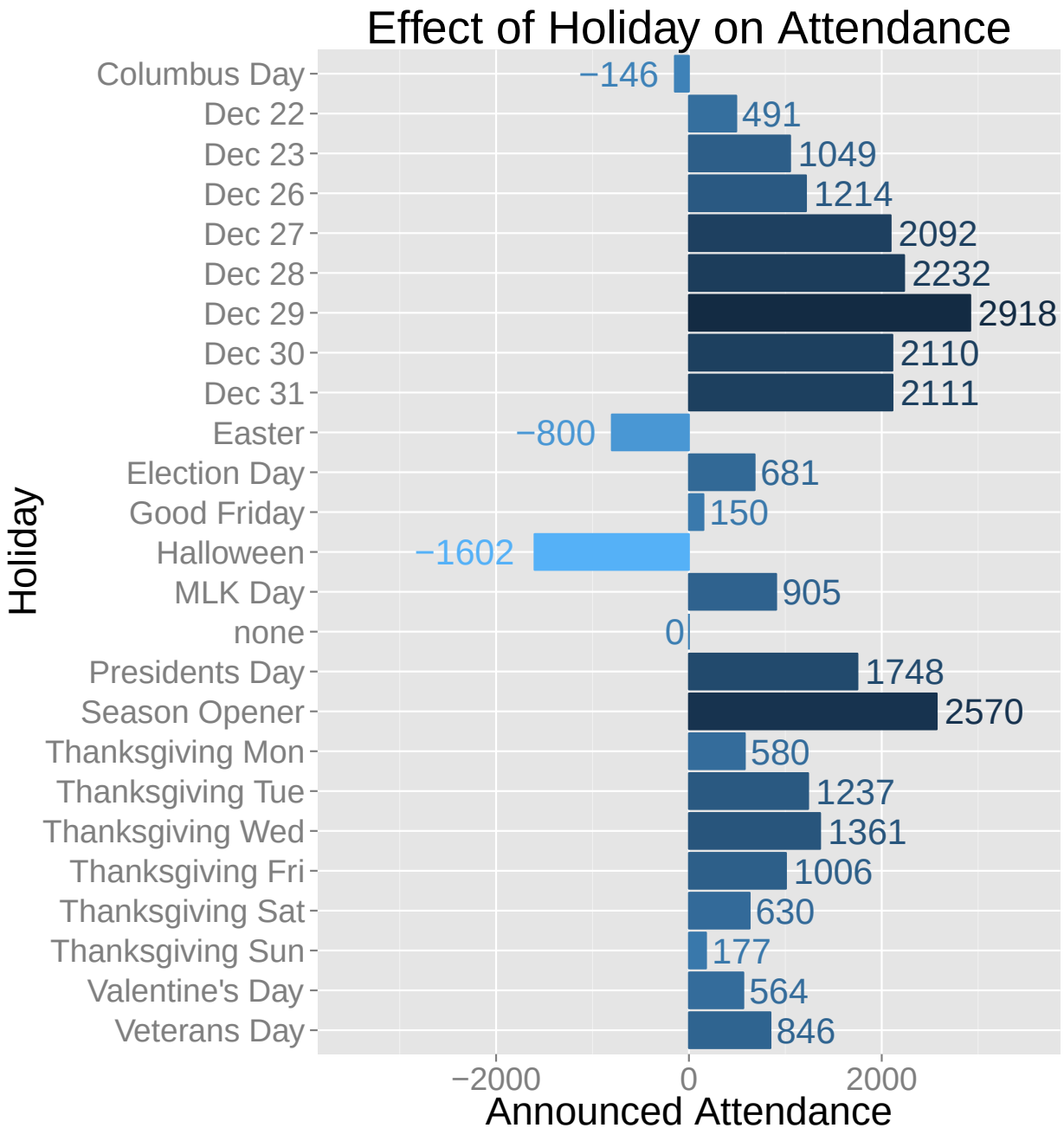


- Attendance increases over the course of the season

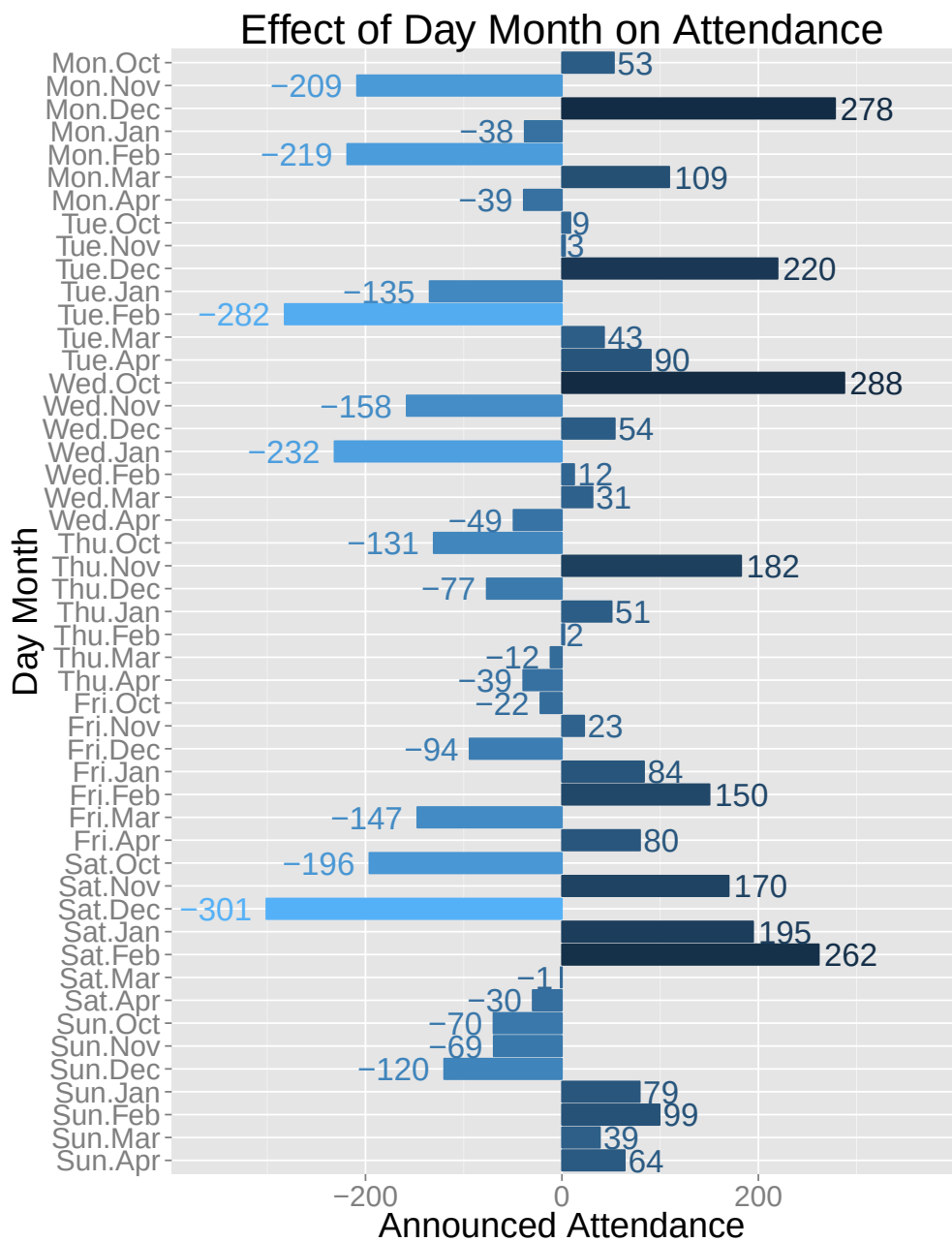
Away Team


Effect of Away on Attendance
DET 1270
NYR 1162
PIT 1154
BOS 821
CHI 618
MTL 583
TOR 360
PHI 307
WSH 273
N.J 88
VAN 54
S.J −46
L.A −52
BUF −52
T.B −69
FLA −107
COL −112
WPG −121
NYI −138
EDM −162
ANA −171
CGY −340
ATL −376
DAL −527
MIN −564
CBJ −592
CAR −614
PHX −623
NSH −660
OTT −669
STL −696
Away
−1000
0
1000
Announced Attendance

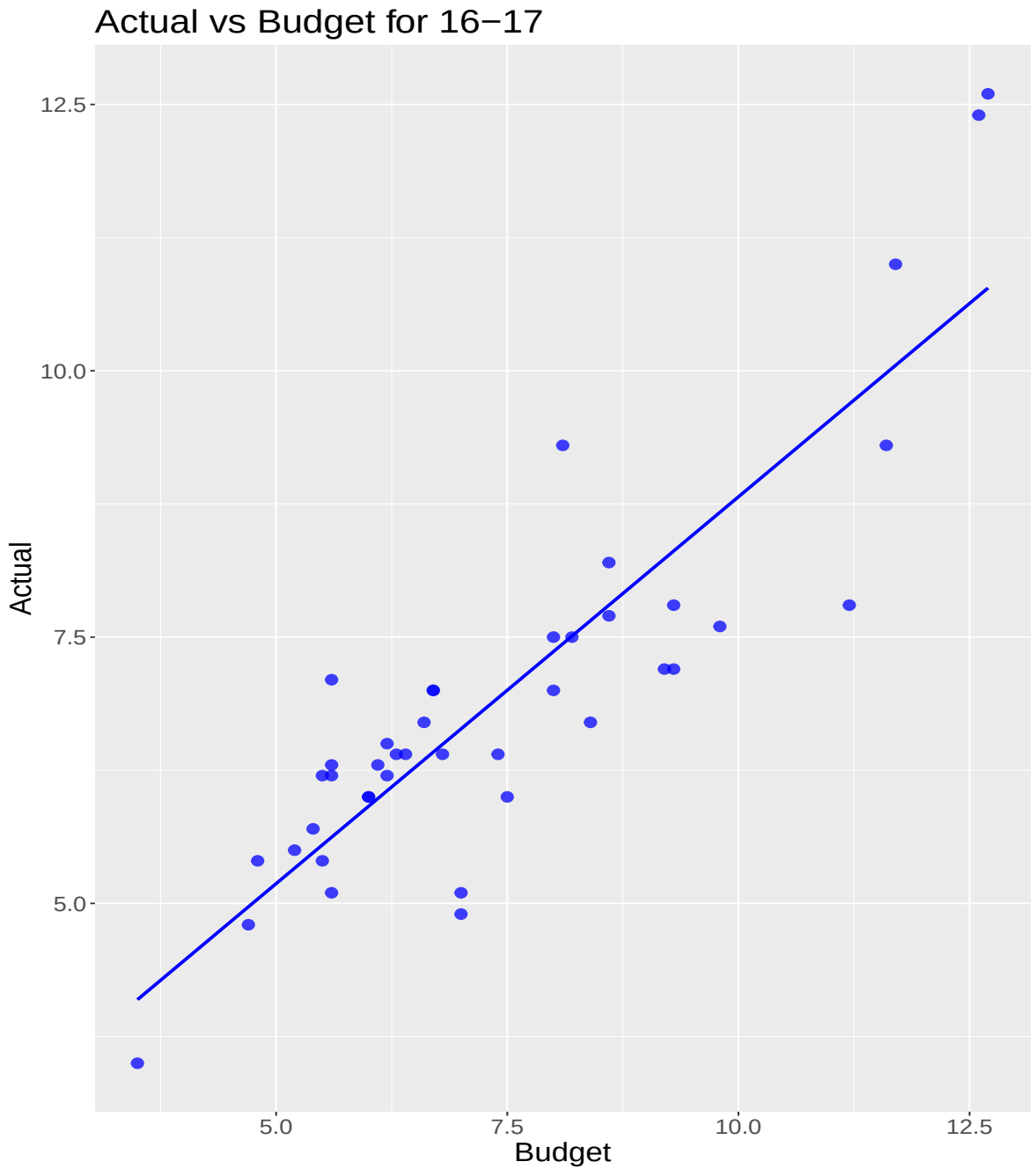
Holidays


Effect of Holiday on Attendance
Holiday
Columbus Day
Dec 22
Dec 23
Dec 26
Dec 27
Dec 28
Dec 29
Dec 30
Dec 31
Easter
Election Day
Good Friday
Halloween
MLK Day
none
Presidents Day
Season Opener
Thanksgiving Mon
Thanksgiving Tue
Thanksgiving Wed
Thanksgiving Fri
Thanksgiving Sat
Thanksgiving Sun
Valentine's Day
Veterans Day
−146
491
1049
1214
2092
2232
2918
2110
2111
−800
681
150
−1602
905
0
1748
2570
580
1237
1361
1006
630
177
564
846
−2000
0
2000
Announced Attendance

Day-month combinations

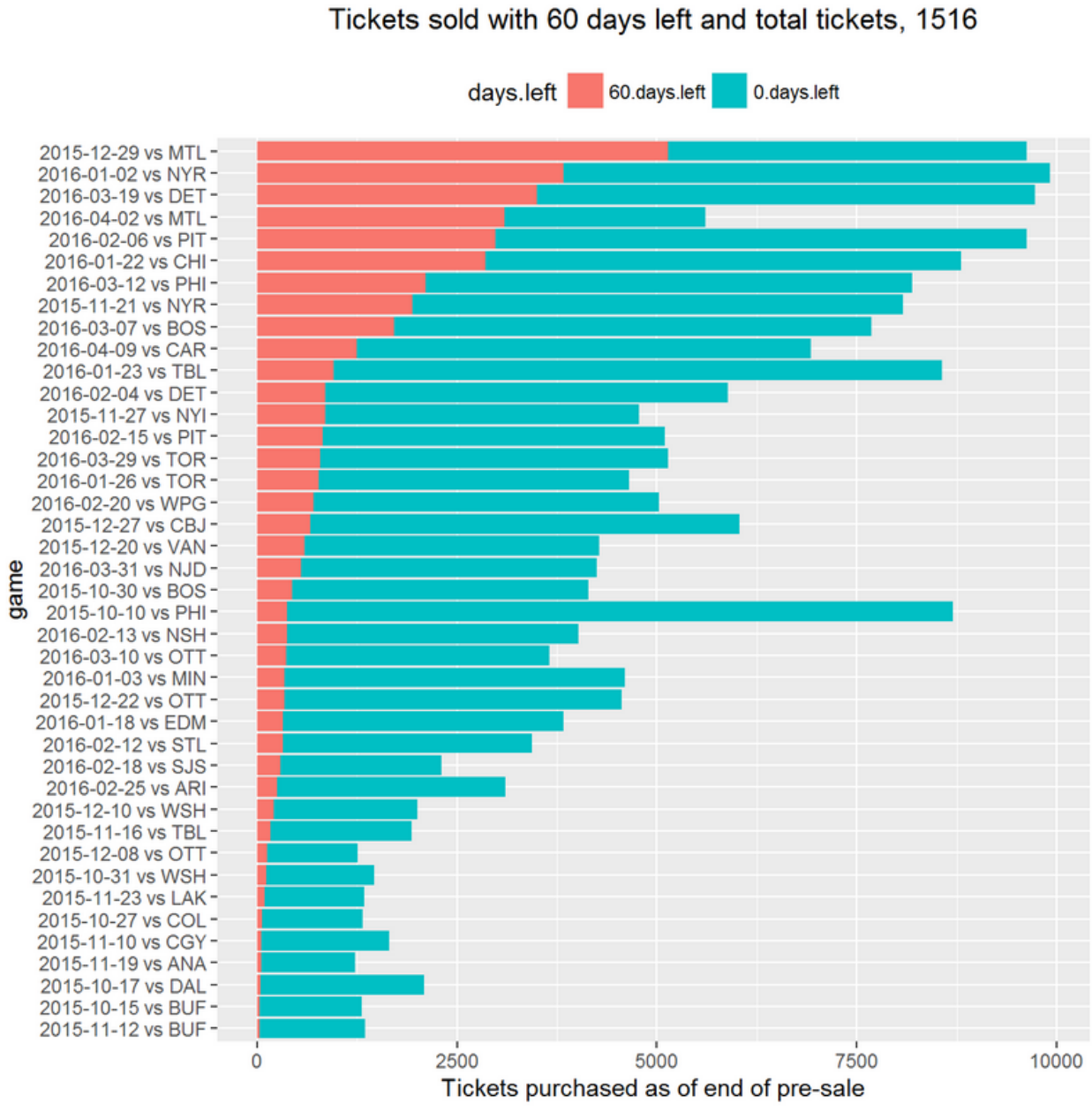


Actual vs Budget for 16-17



Total tickets and tickets 60 days out

Total tickets and tickets 60 days out



What are students exposed to?

Ideas I focus on when giving this talk to students.

- Data exploration/visualization
- Multivariate thinking, need for regression or something else
- Interpretation of regression coefficients
- Solving real business problem

What are students exposed to?

Ideas I focus on when giving this talk to students.

- Data exploration/visualization
- Multivariate thinking, need for regression or something else
- Interpretation of regression coefficients
- Solving real business problem
- Validating a model
  - Don't discuss cross-validation, out-of-sample testing, regression diagnostics
  - Do discuss real-life validation

What are students exposed to?

Ideas I focus on when giving this talk to students.

- Data exploration/visualization
- Multivariate thinking, need for regression or something else
- Interpretation of regression coefficients
- Solving real business problem
- Validating a model
  - Don't discuss cross-validation, out-of-sample testing, regression diagnostics
  - Do discuss real-life validation
- Projects that have multiple applications
  - Think big, not small
  - Small questions often require a bigger model
  - Benefit: can answer other questions with the same model

What are students exposed to?

Ideas I focus on when giving this talk to students.

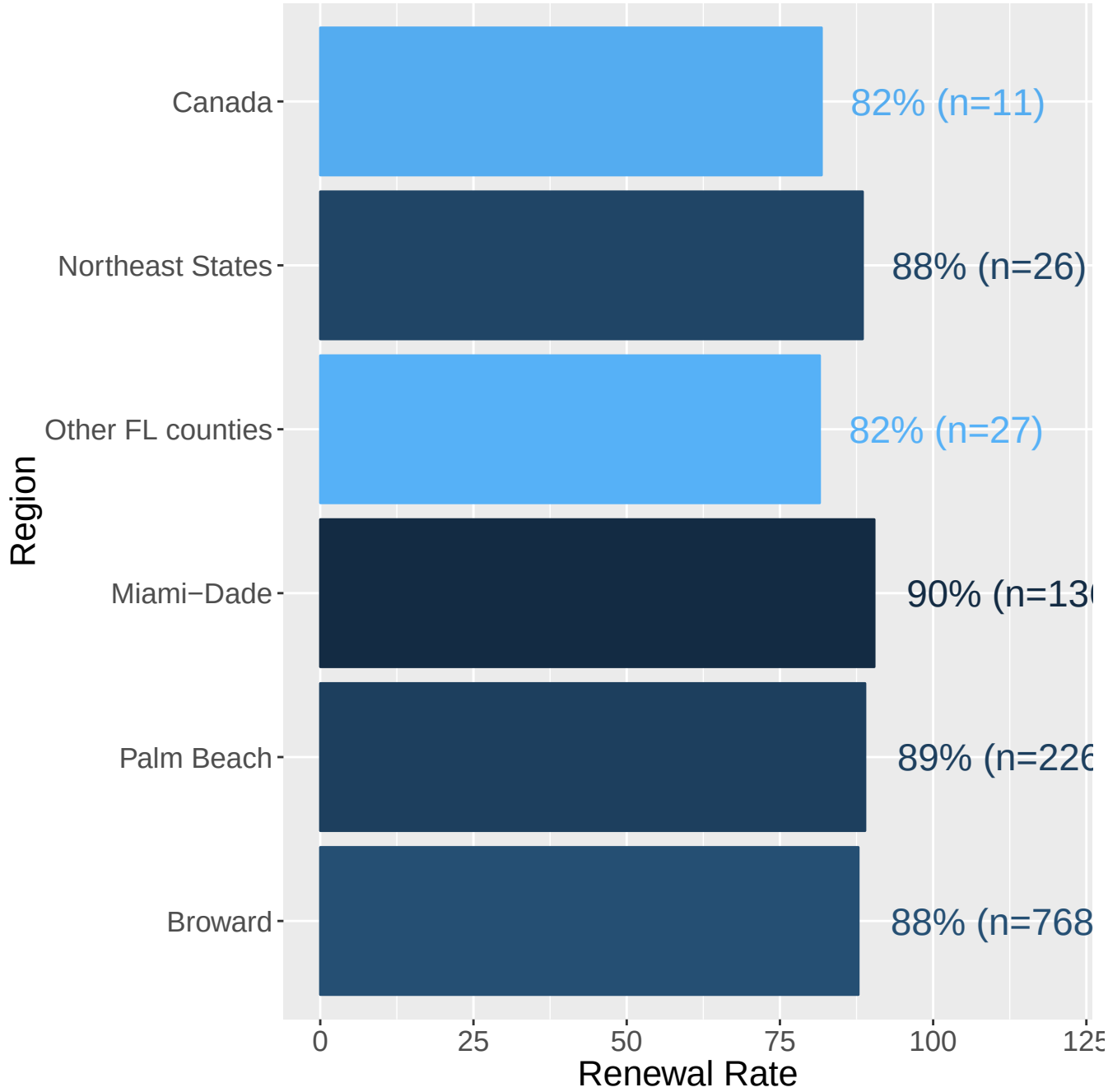
- Data exploration/visualization
- Multivariate thinking, need for regression or something else
- Interpretation of regression coefficients
- Solving real business problem
- Validating a model
  - Don't discuss cross-validation, out-of-sample testing, regression diagnostics
  - Do discuss real-life validation
- Projects that have multiple applications
  - Think big, not small
  - Small questions often require a bigger model
  - Benefit: can answer other questions with the same model

What I don't focus on

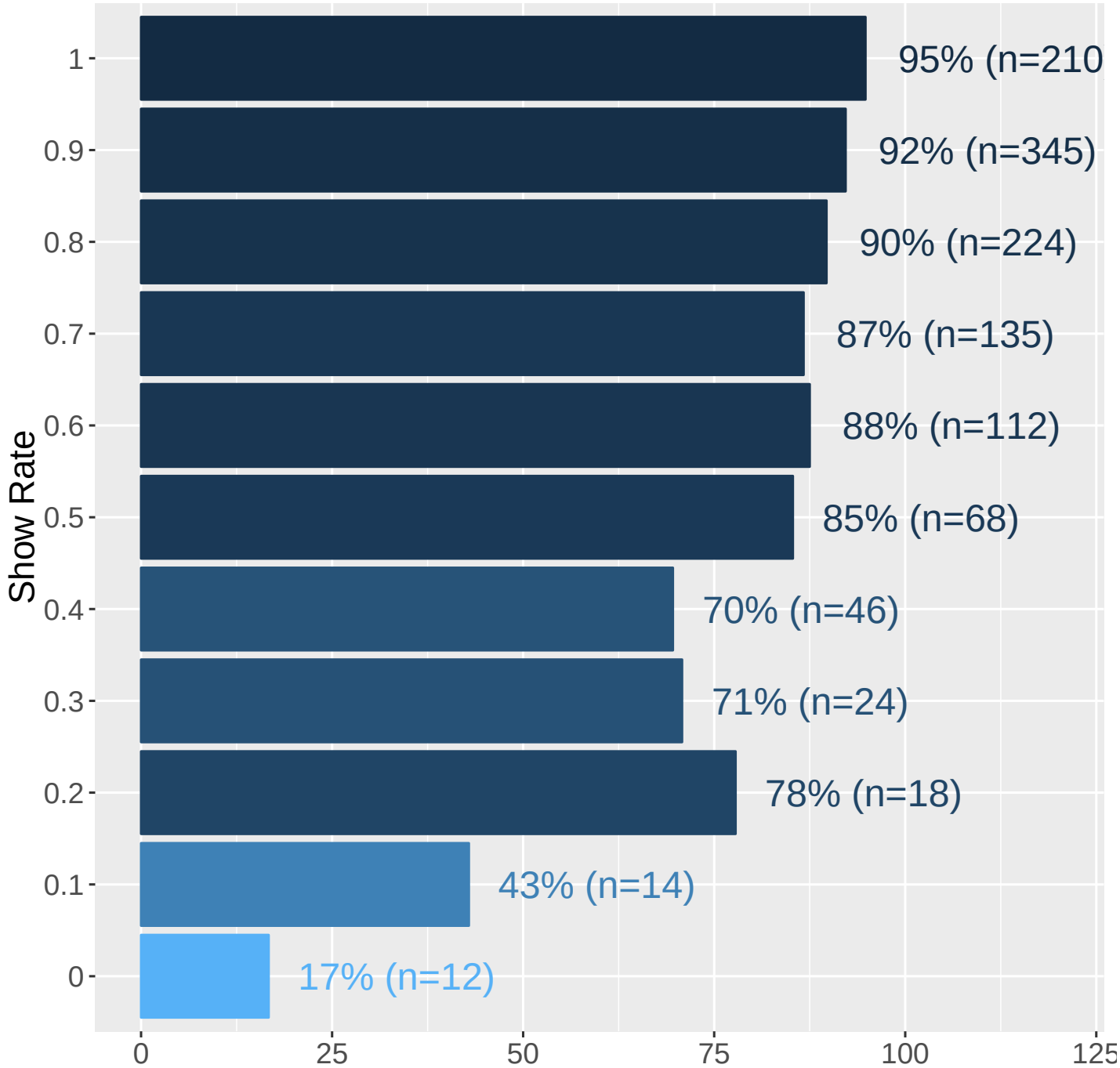
- Data acquisition, clean, reorganizing, merging

Renewal Prediction

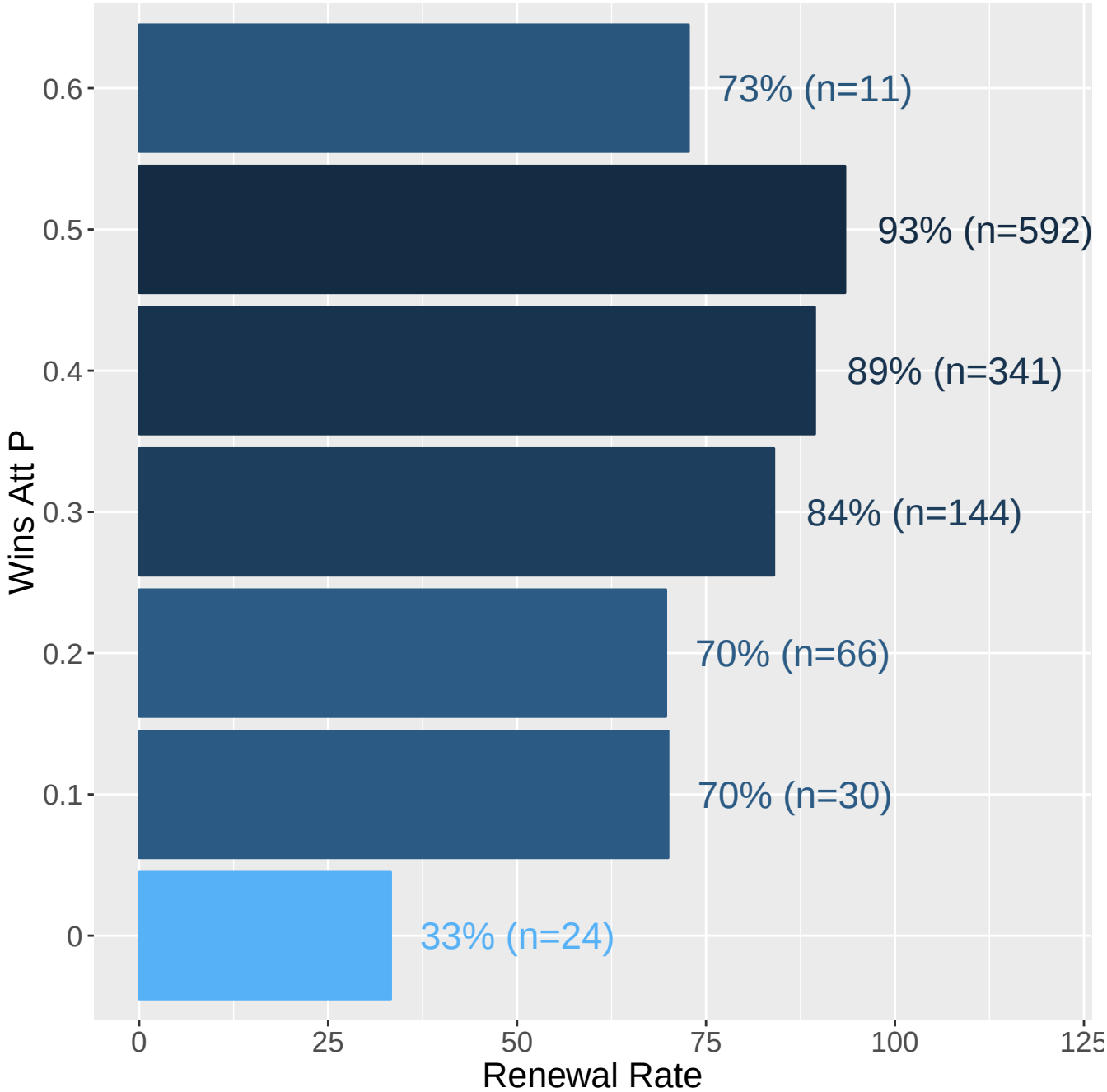
Renewal Prediction



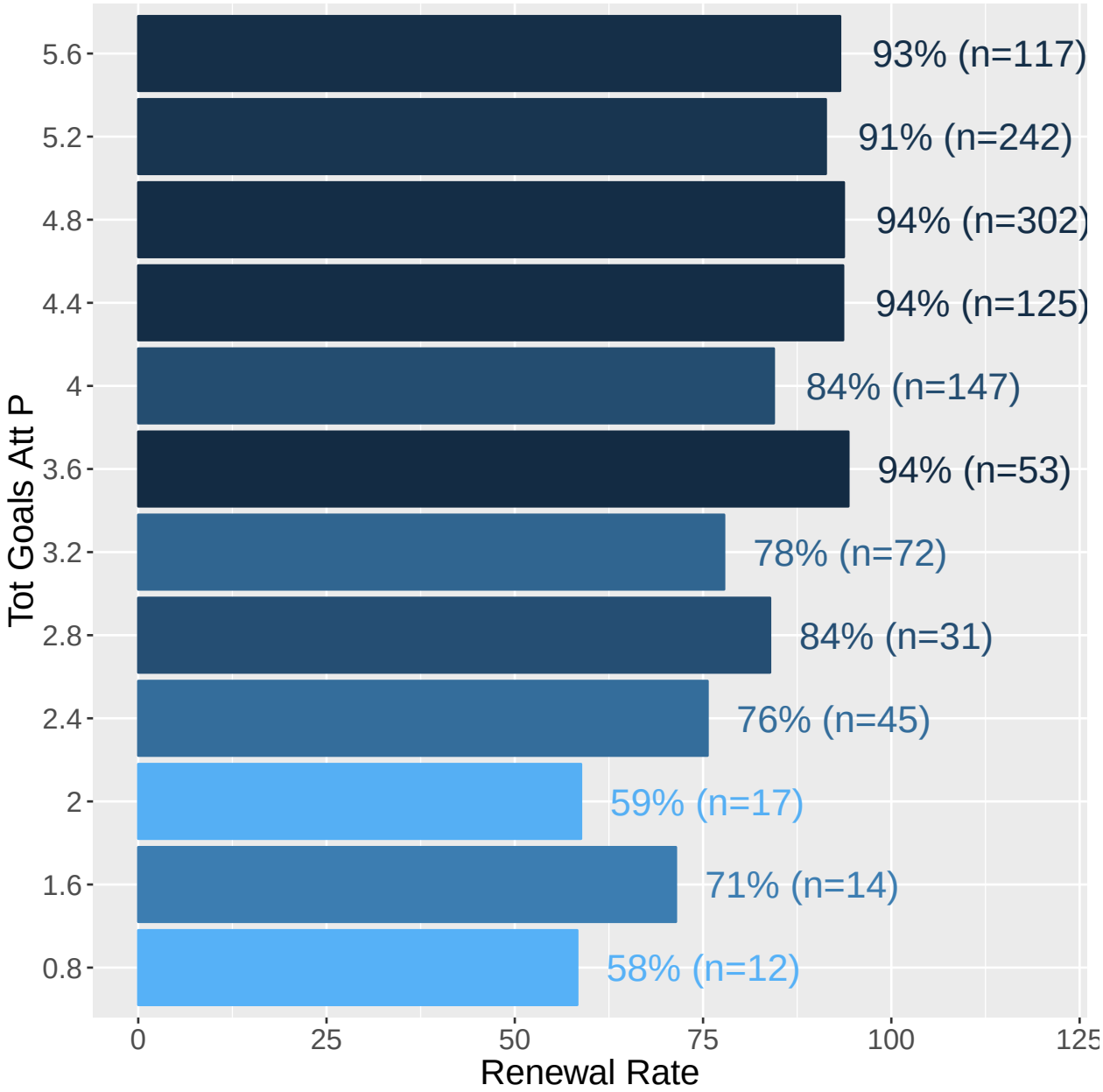
Show rate


Show Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
95% (n=210
92% (n=345)
90% (n=224)
87% (n=135)
88% (n=112)
85% (n=68)
70% (n=46)
71% (n=24)
78% (n=18)
43% (n=14)
17% (n=12)
0
25
50
75
100
125

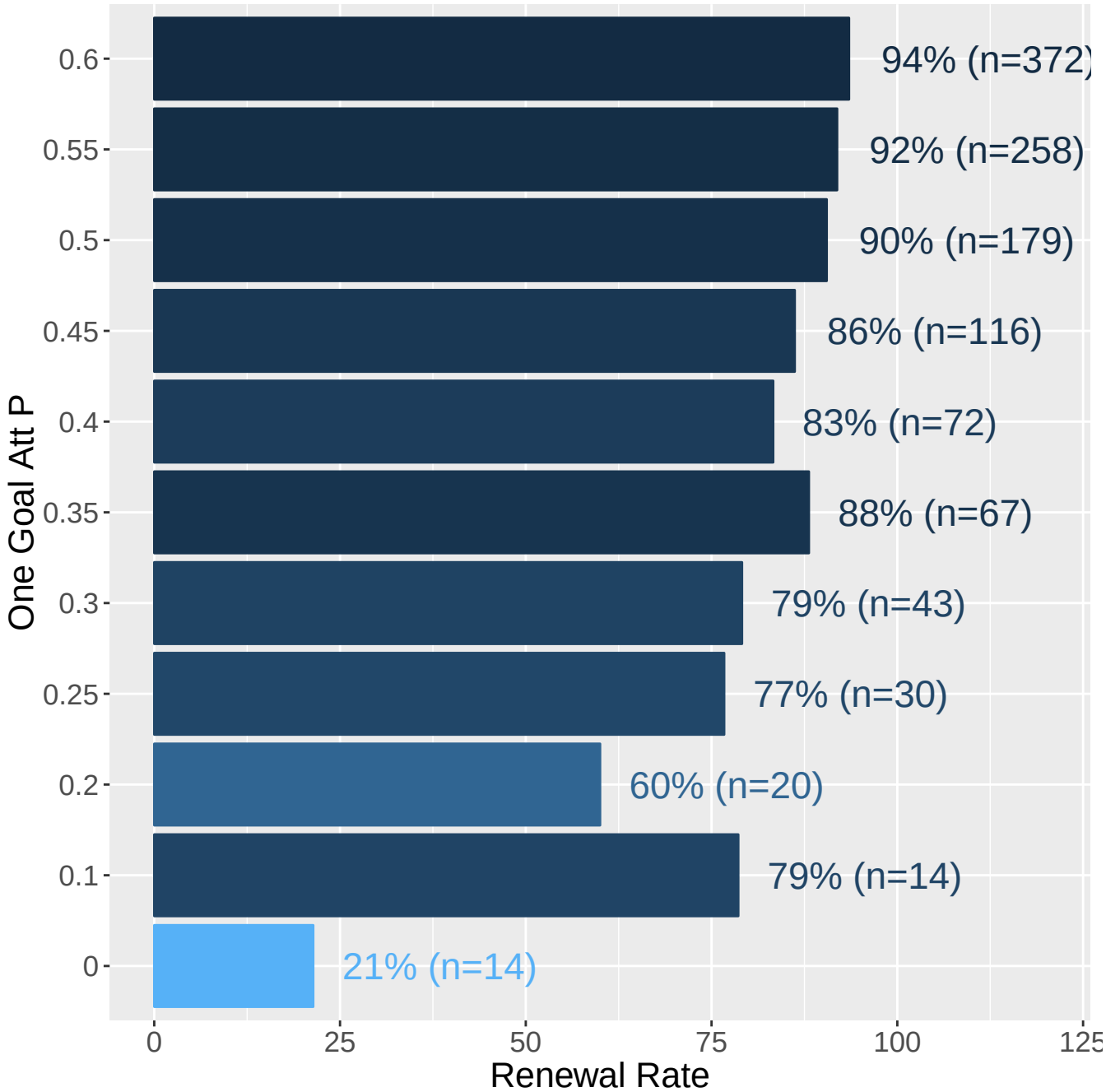
Win% in games attended



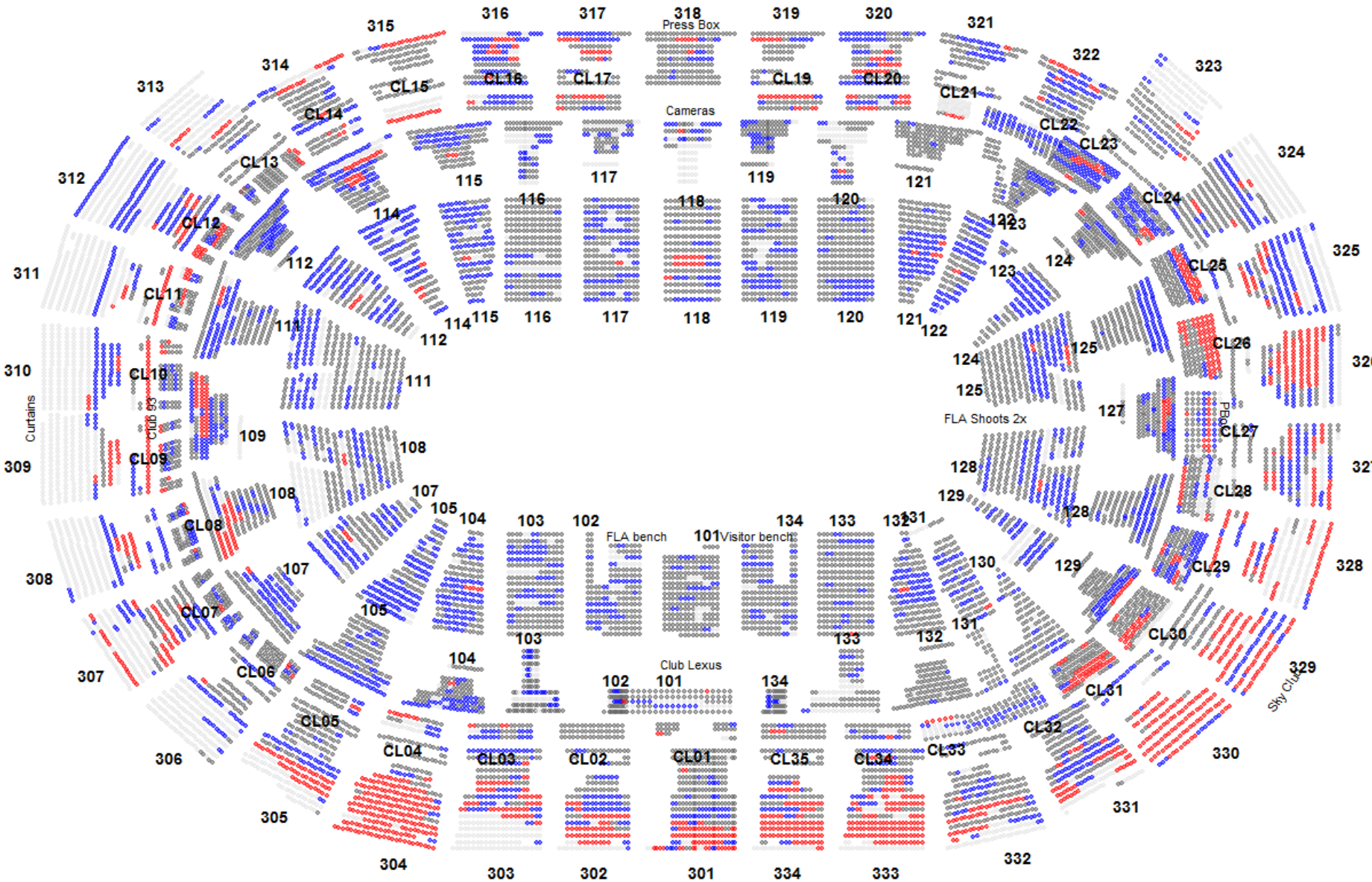
Average total goals in games attended



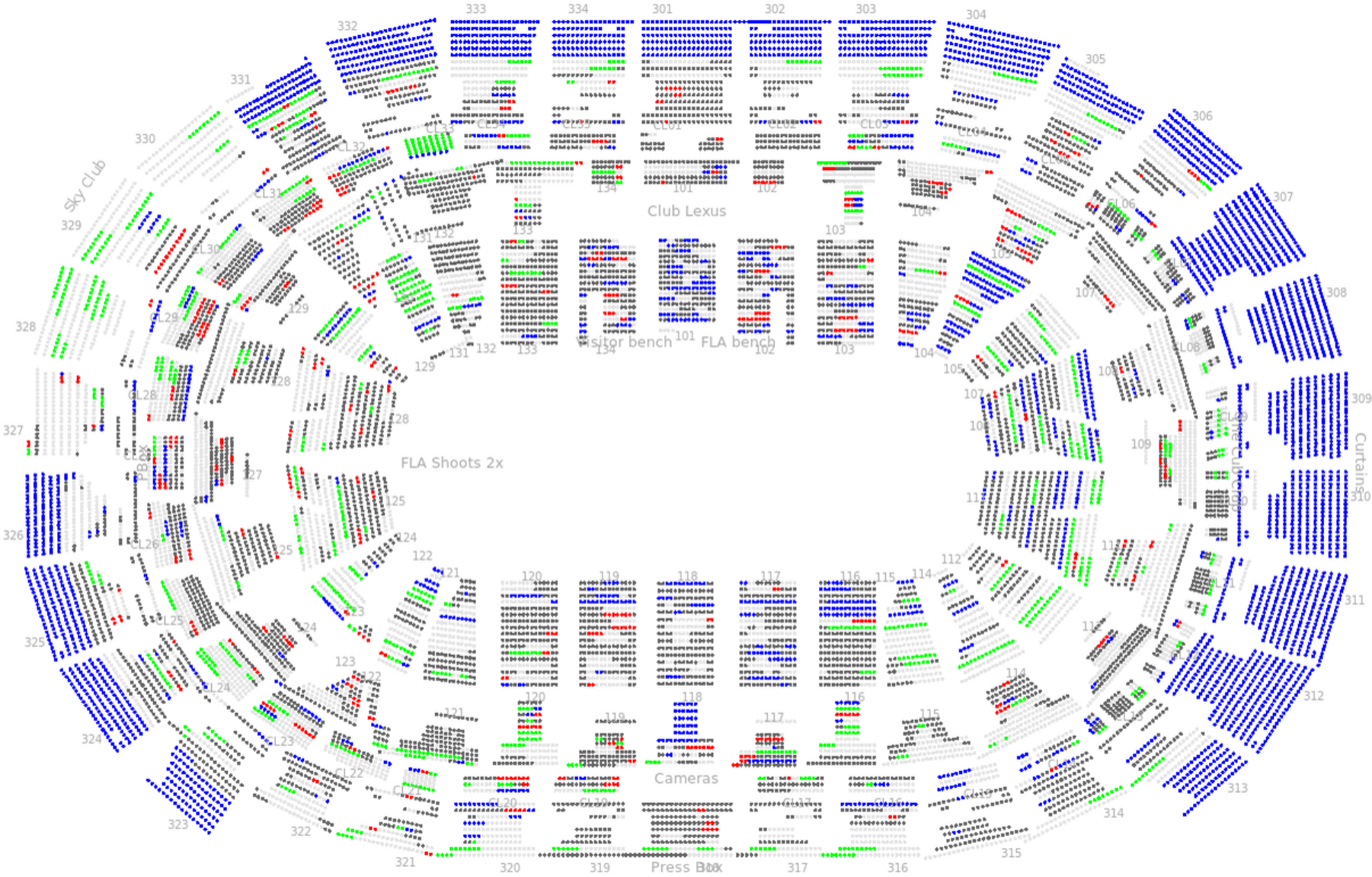
Proportion of 1-goal games in games attended



Data visualization



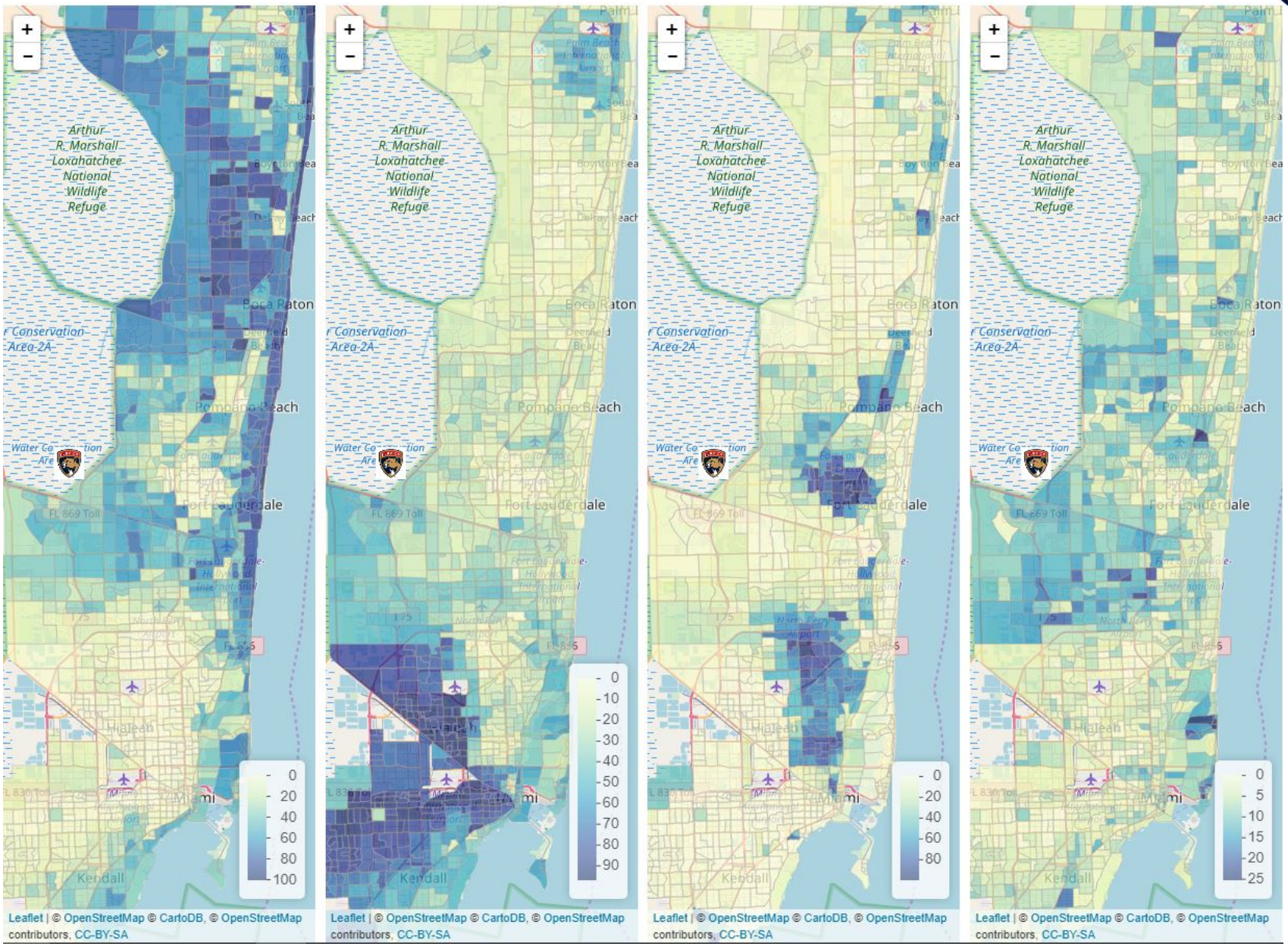
Arena maps


Sky Club
Club Lexus
Visitor bench
FLA bench
FLA Shoots 2x
Cameras
Press Box
Curtains

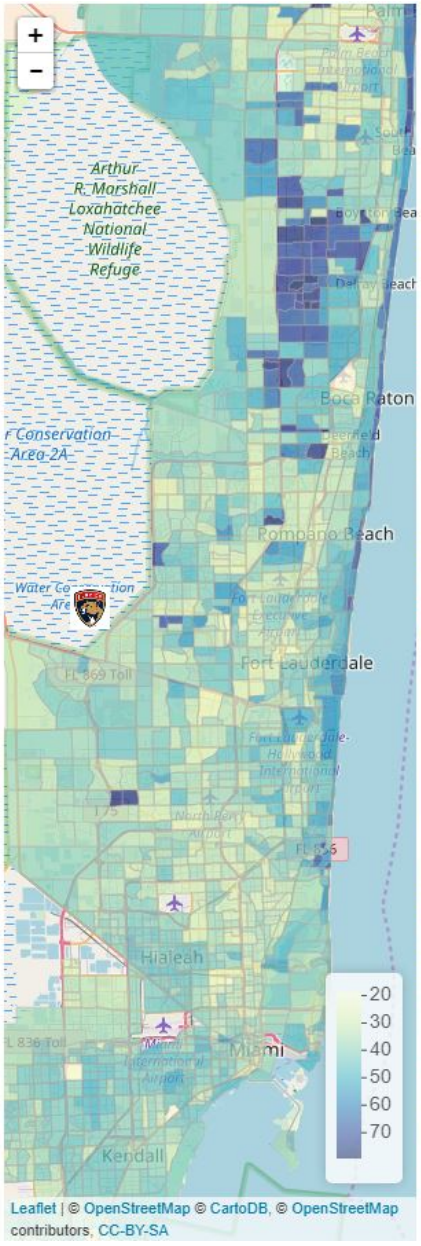
Note to Self

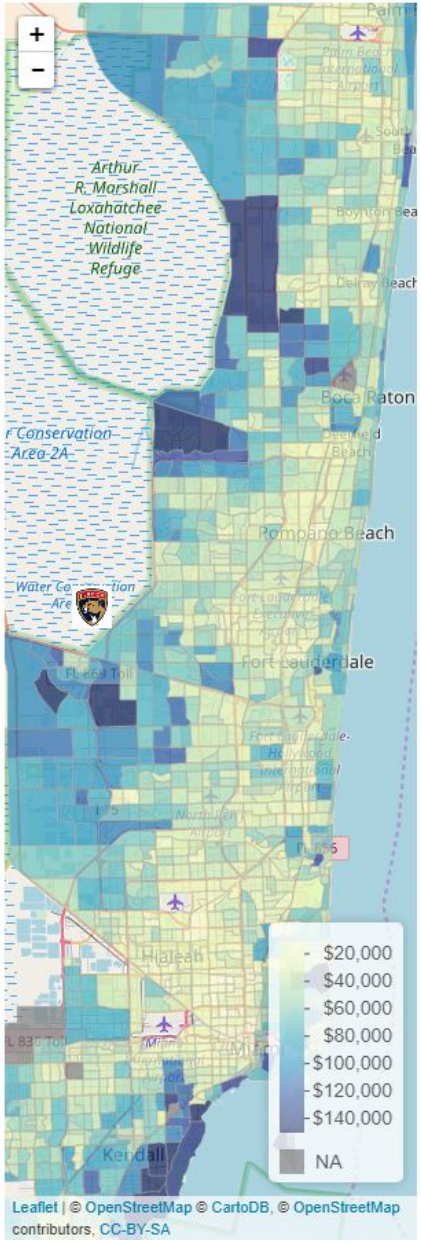
Back to other slides!

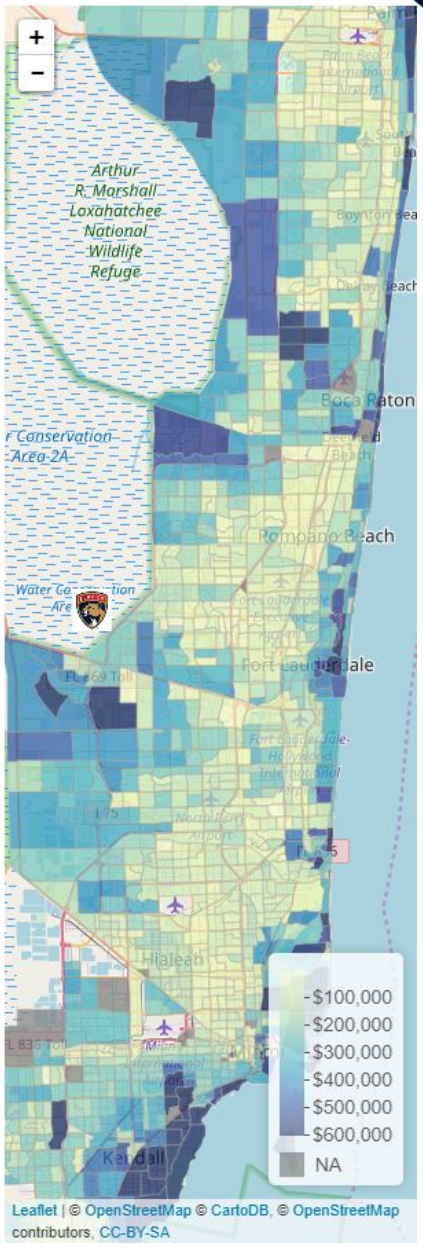
South FL maps



South FL maps







Overview of Hockey Data

Kinds of data

- Play by play
- Shifts
- Shot locations

Overview of Hockey Data

Kinds of data

- Play by play
- Shifts
- Shot locations

What can be done with this data?

Overview of Hockey Data

Kinds of data

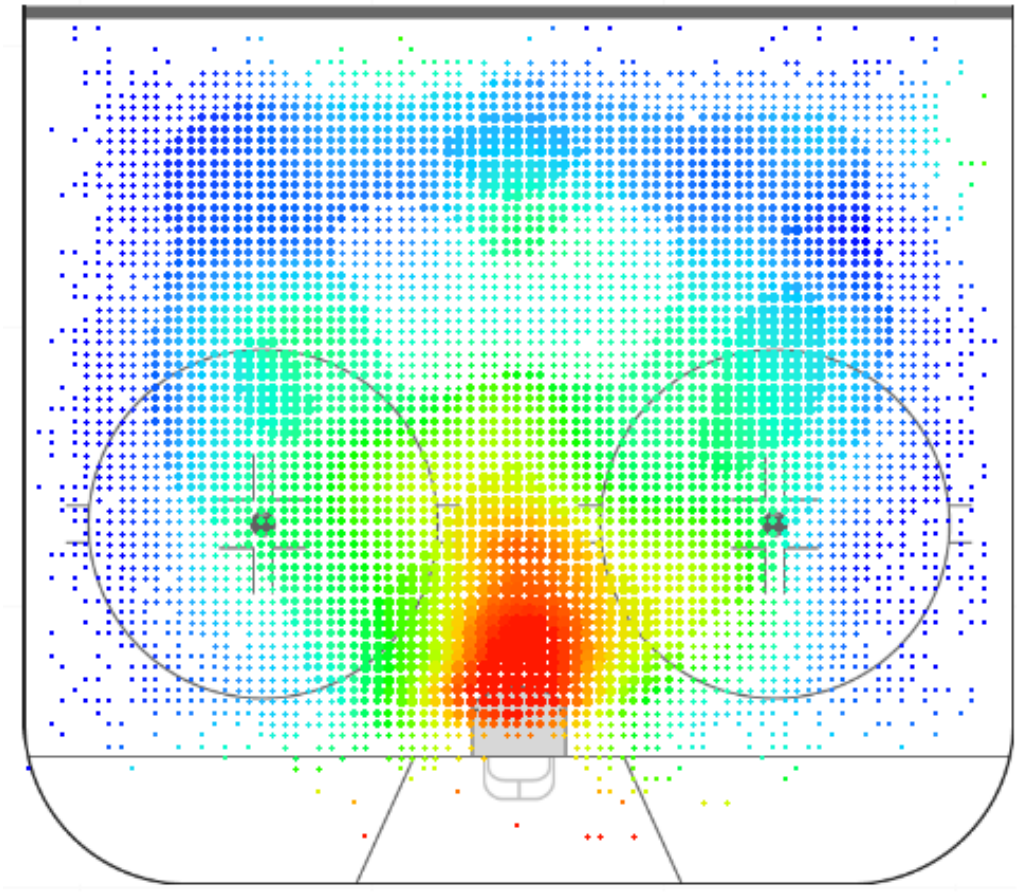
- Play by play
- Shifts
- Shot locations

What can be done with this data?

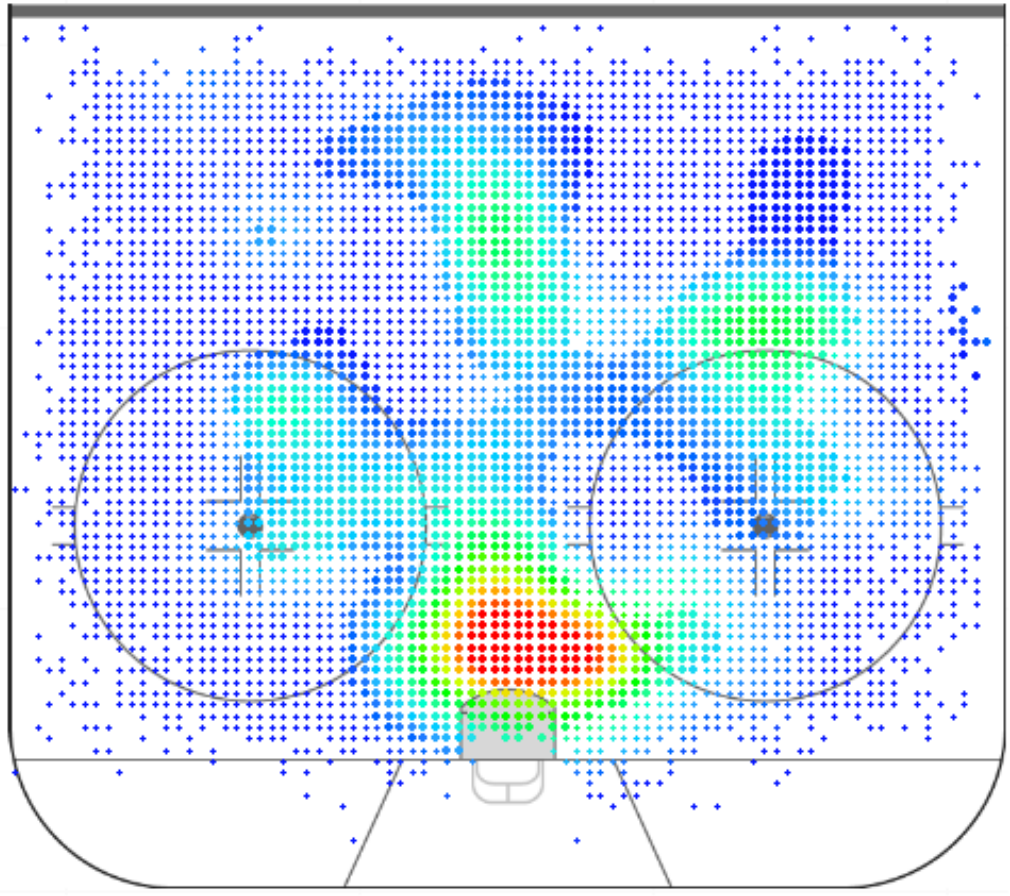
- Player performance metrics
- Team performance metrics
- Data visualization

Data visualization: League data

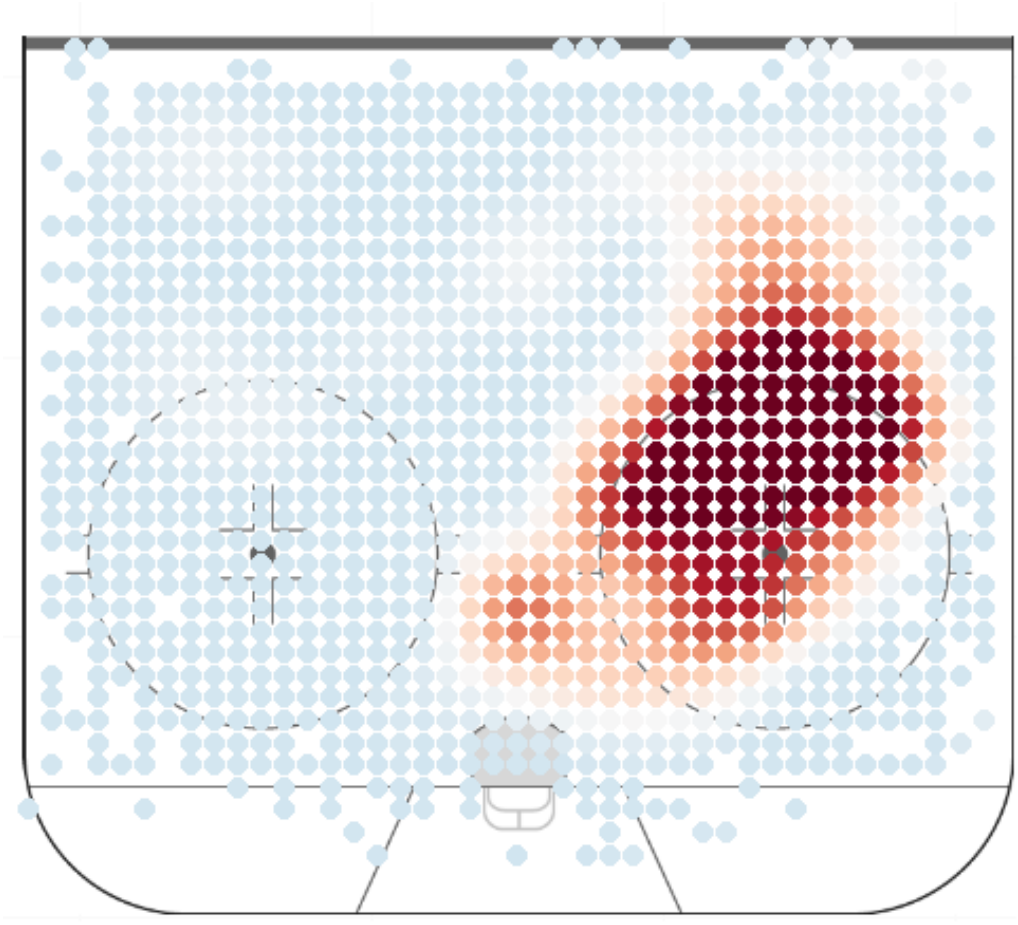
5v4



Data visualization: Team data



Data visualization: Player data



Data visualization: Player data