Homework_4_wi19b004

Michael Scheidl

03 10 2021

Contents

1 Assignment

The MLB teams data set in the mdsr package contains information about Major League Baseball teams in the past four seasons. There are several quantitative and a few categorical variables present. See how many variables you can illustrate on a single plot in R . The current record is 7. (Note: This is not good graphical practice— it is merely an exercise to help you understand how to use visual cues and aesthetics!)

2 Solution

2.1 Libraries

```
library(tidyverse)
library(mdsr)
```

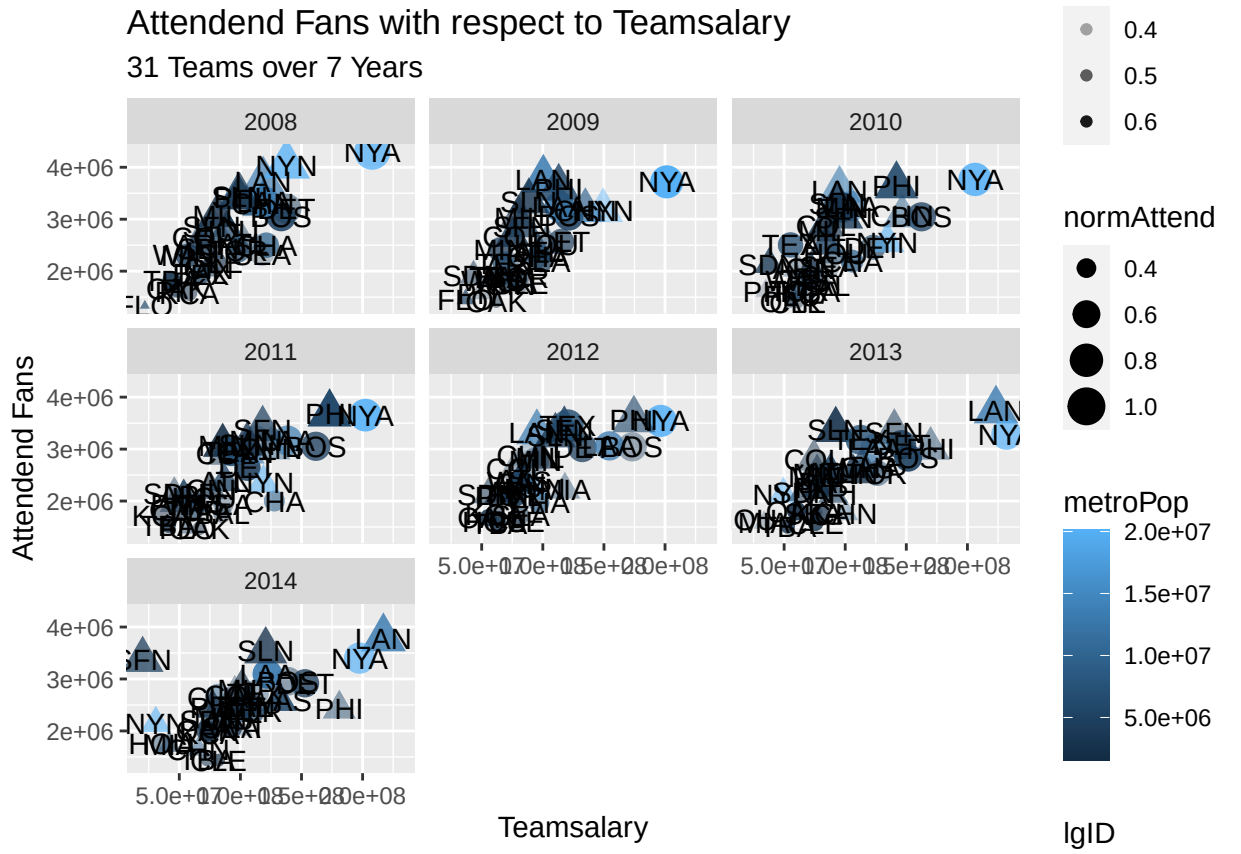
2.2 Inspect Data

```
#?MLB_teams
head(MLB_teams, 1) %>% as_tibble()
```

```
## # A tibble: 1 x 11
##   yearID teamID lgID      W     L  WPct attendance normAttend  payroll metroPop
##    <int> <chr>  <fct> <int> <int> <dbl>      <int>      <dbl>    <int>    <dbl>
## 1   2008 ARI    NL       82    80 0.506    2509924      0.584 66202712  4489109
## # ... with 1 more variable: name <chr>
```

2.3 Plot

```
g <- MLB_teams %>% as_tibble() %>%
# Put data in ggplot2 with Payroll on the x and attendace on the y axis
  ggplot(aes(x=payroll,y=attendance)) +
# Use normalized Attendance as Size, HomeTowns population as Color
  geom_point(aes(size=normAttend, col = metroPop,
# Leage as Shape and WinningPercentage as Alpha Value
                 shape=lgID, alpha=WPct)) +
  facet_wrap(~yearID) + # Split Plot per Season
  geom_text(aes(label = teamID)) + # Add Team Name as Label
  xlab("Teamsalary") + # Label X Axis
  ylab("Attendend Fans") + # Label Y Axis
  labs(
# Set a Meaningfull Title and subtitle to describe Plot and used Dataset
     title = "Attendend Fans with respect to Teamsalary",
     subtitle = paste(nlevels(as_factor(MLB_teams$teamID)), "Teams over",
                      nlevels(as_factor(MLB_teams$yearID)), "Years"))
# NOTE: Could not resize the plotting Area to show complete Legend and Texts
g # Show Plot
```



Illustrated 7 Variables in one (very small) Plot and tied the record.