

Homework 04

DSO 545: Statistical Computing and Data Visualization

Fall 2018

Instructions

- Use R Markdown to answer all question. Write R code for each question.
- Submit your R Markdown file as well as your (pdf or word files) to blackboard with comments when needed
- USC won't tolerate any kind of cheating
- Good luck!

Case 1: English Premiere League

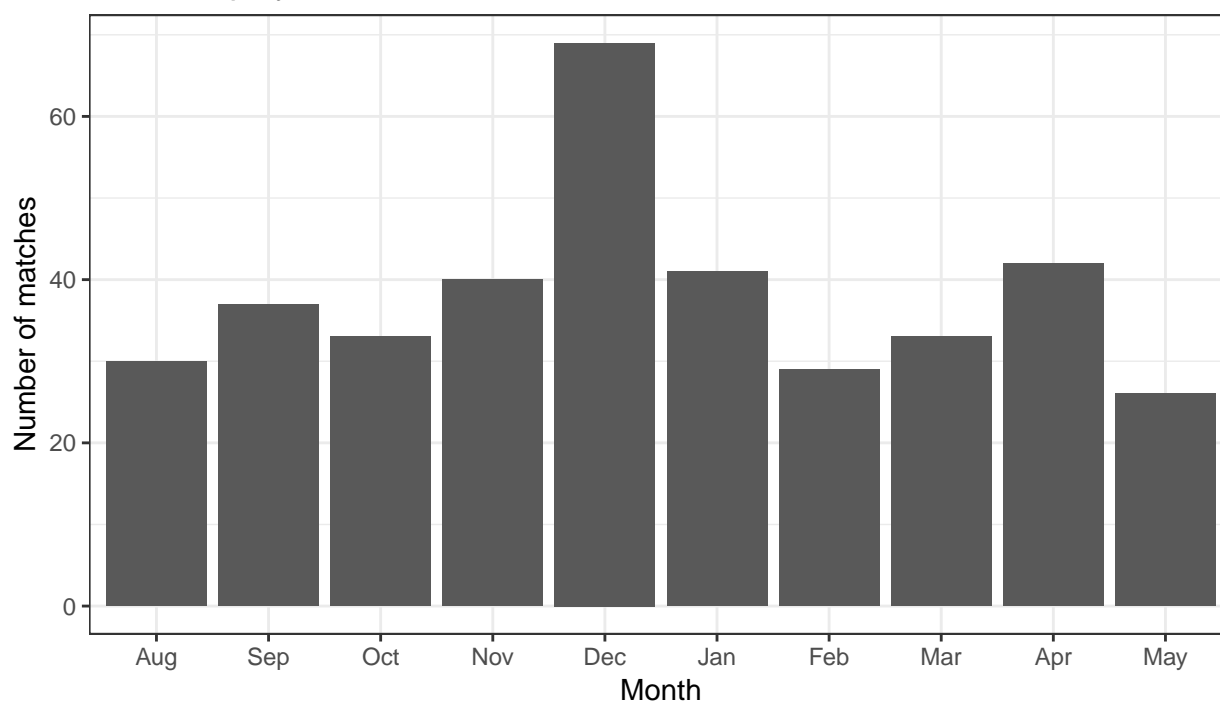
This `ep1.csv` dataset contains results from every Premier League match from 1993-1994 to 2017-2018.

Metadeta

Variables	Description
HomeTeam	The name of the home team
AwayTeam	The name of the away team
FTHG	The total number of goals scored by the home team during the match at full time
FTAG	The total number of goals scored by the away team during the match at half time
FTR	The full time result, denoted as 'H' for home team win, 'A' for away team win, or 'D' for draw
Season	The season in which the match was played
Match_Date	The date the match was played
Home_Win	If Home Team Won "1", else 0
Away_Win	If Away Team Won "1", else 0
Draw	If match is a draw "1", else 0

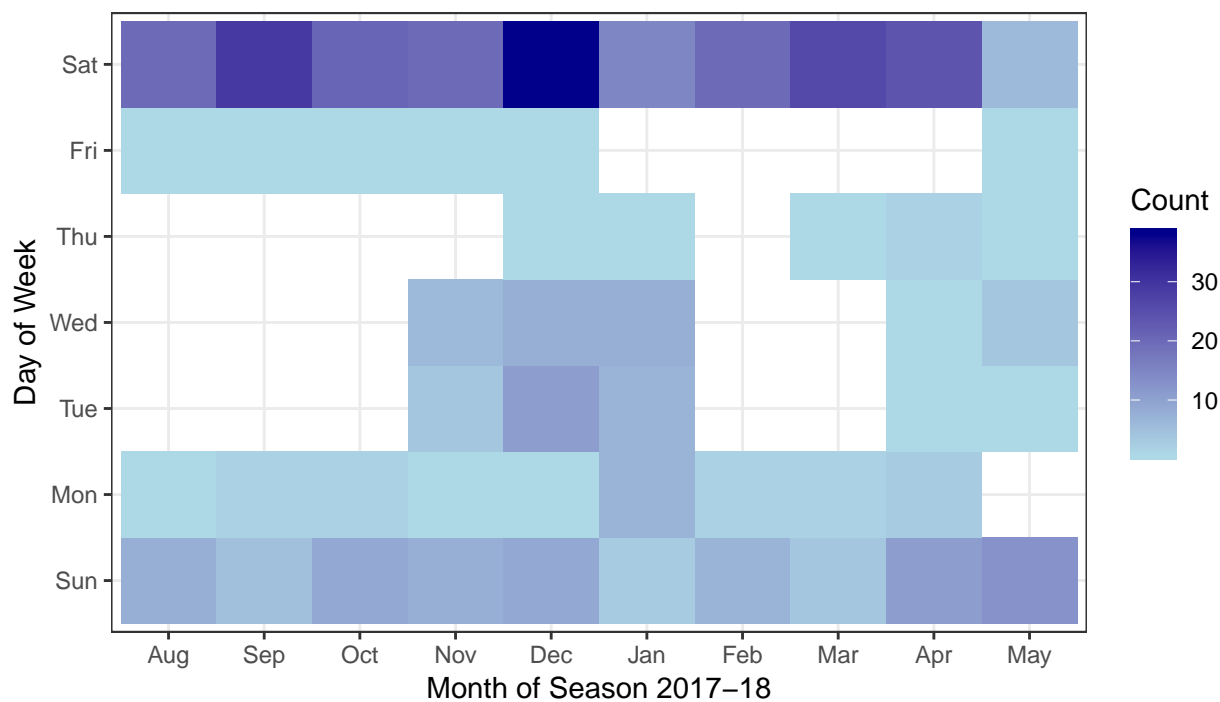
1. **(3 points)** How many matches did Liverpool win from 2001 (inclusive) to 2010 (inclusive)?
2. **(3 points)** Find out the top 5 years in which the highest number of matches were played?
3. **(3 points)** What is the number of games played in each month in the year 2015.
4. **(3 points)** Which month in 2017 had the highest number of matches played in a day?
5. **(3 points)** In 2017-18 season, what is the number of matches played on each day of the week?
6. **(4 points)** The following plot shows the number of matches played in each month in the 2017-18 season. Replicate it in R.(Hint: Use `theme_bw()`)

Matches played in the 2017–2018 Season



7. (4 points) The following plot shows the number of matches played in a week in the 2017-18 season. Replicate it in R. (Hint: Use `theme_bw()`, and colors ranging from “lightblue” to “darkblue”)

Matches played in 2017–18 season by Month and Day of Week



Case 2: Starbucks across the US

The dataset `starbucks_US_count.csv` provides information on the number of Starbucks stores across US.

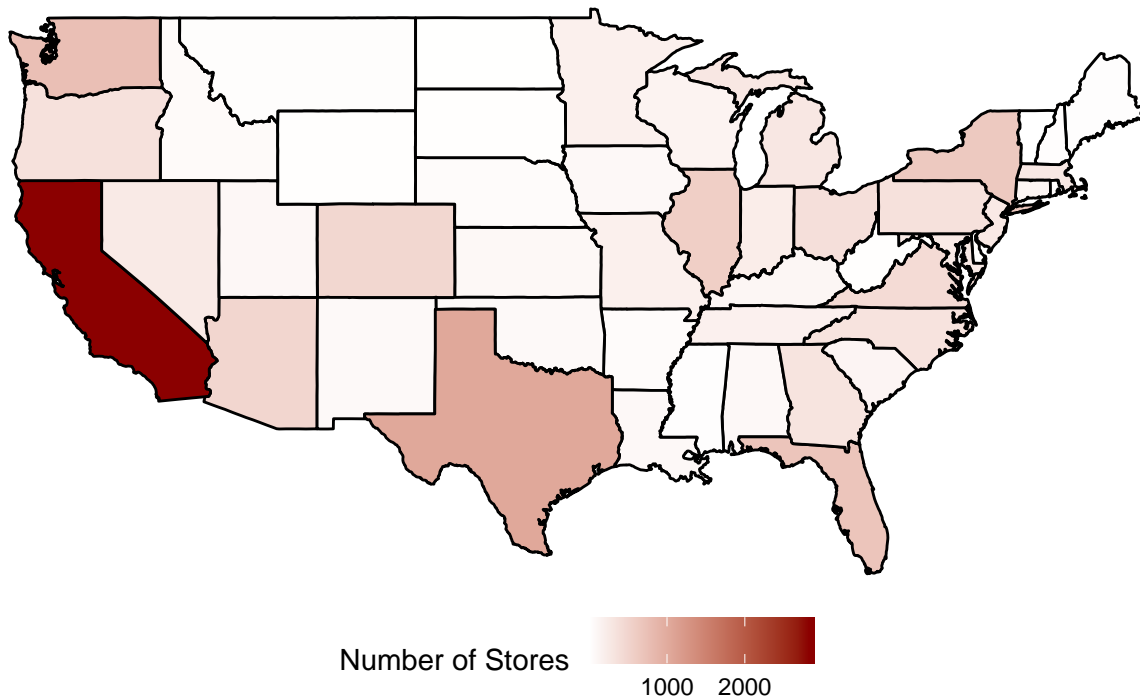
Metadata

Variables	Description
Brand	Brand
Store.Number	Store Number
Store.Name	Store Name
Ownership.Type	Ownership Type
Street.Address	Street Address
City	City
State.Province	State/Province
Country	Country
Postcode	ZIP
Phone.Number	Phone Number
Timezone	Timezone

8. (4 points) Create an **EXACT** copy of the following graph. Dataset used is “starbucks_US”. (Hint: Check out the function `theme()` in `ggplot` in order to have the map legend/guide at the bottom)

Specifications: Gradient of *white* and *darkred*, plot title size is *16 bold* and legend title size is *10 bold*.

Starbucks Franchise across United States



Case 3: Storytelling with Data

9. **(3 points)** Read Chapter 3 from “Storytelling with data: A Data Visualization Guide for Business Professionals”, and write a **max of 1 page summary** of the key points in the chapter. Please note, that this book is available online on USC’s library webpage.