

# Homework 03

*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!

## Part 1: Congress Age

This dataset has an entry for every member of congress who served at any point during a particular congress between January 1947 and February 2014. The age of the representatives at the time of service has been captured in this dataset.

The United States Congress is the bicameral legislature of the Federal government of the United States. The legislature consists of two chambers: the Senate and the House of Representatives.

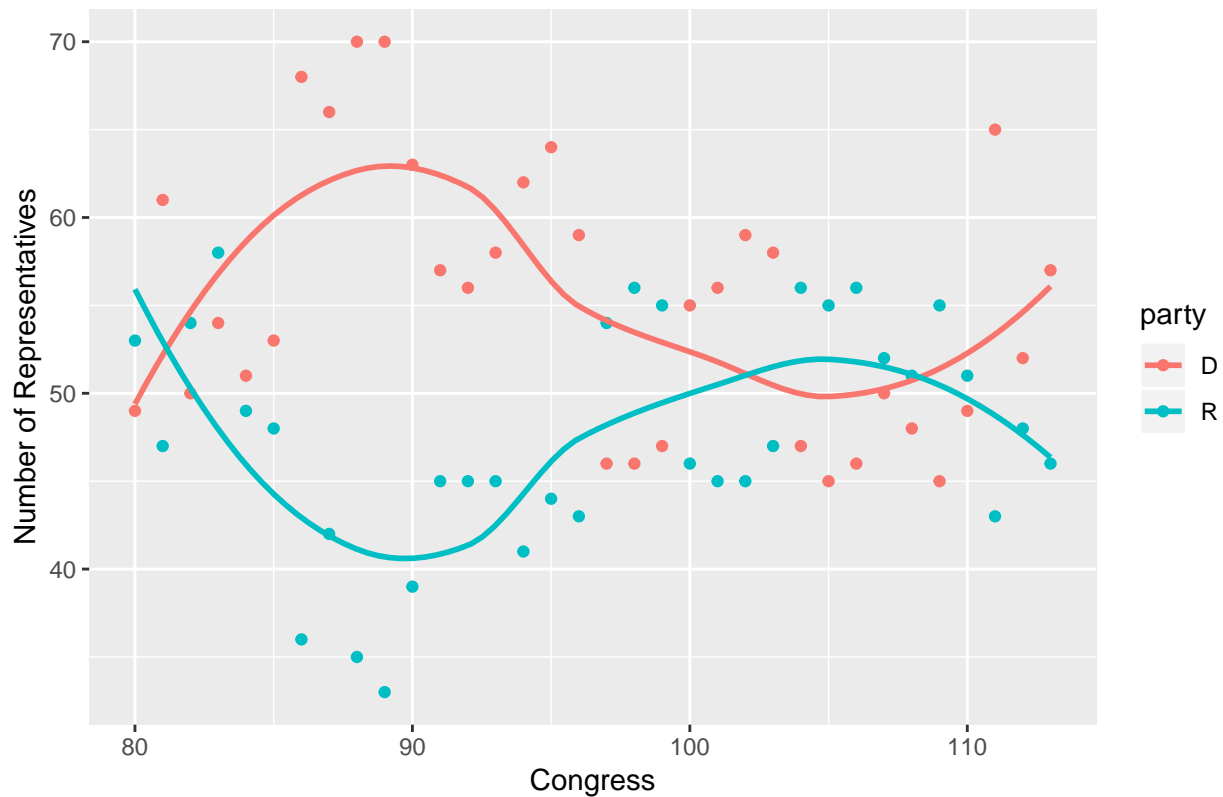
## Variables

Variables	Description
congress	United States Congress
chamber	House or Senate
bioguide	Biographical Directory ID
firstname	First Name
lastname	Last Name
suffix	Suffix
birthday	Date of Birth
state	State
party	Party
incumbent	Incumbent - yes or no (i.e. currently in office or not?)
termstart	Start of Term
age	Age

1. **(1 point)** What is the mean age of congresswo/men (cogresswomen and congressmen) in the latest congress (congress 113)?
2. **(1 point)** For each party, what is the total number of congresswo/men who served since January 1947 (whole dataset)?
3. **(1 point)** For each party in each chamber, what is the total number of congresswo/men who served in Congress 92.
4. **(1 point)** For each party, find the number of congresswo/men who served in the 112<sup>th</sup> **senate**, their maximum age, minimum age and average age.

5. **(2 points)** In the latest congress (Congress 113), which state has the highest number of **democrat house** representatives, and what is the number?
6. **(3 points)** Display the trend of number of **senate** representatives across the congress (80 to 113) for the both democrats and republicans. (Hint: Use `geom_smooth(se = F)` to display the trends).

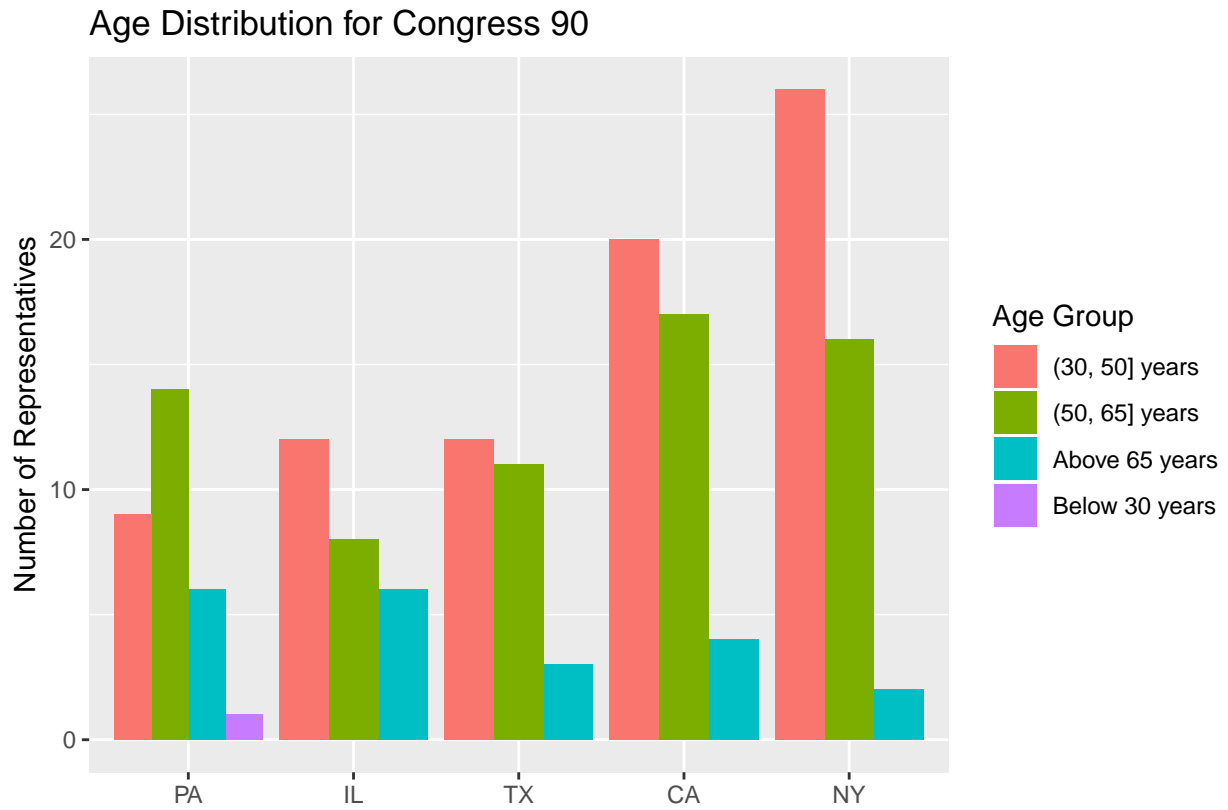
### Trend of Senate Representatives



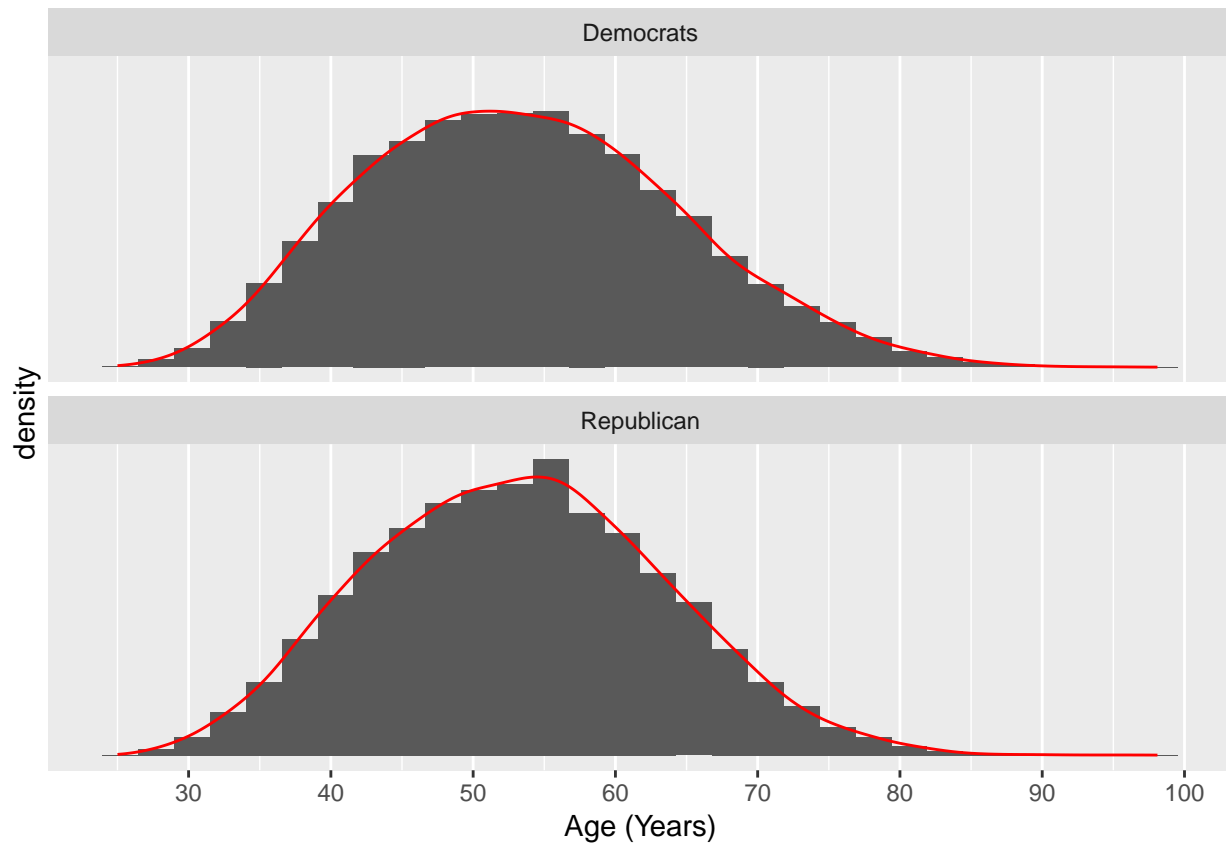
7. **(2 points)** Create a new column **age\_group** according to the below categories (round the age to the nearest integer)

Age	Age.Range
age<=30	Below 30 years
Age>30 and Age<=50	(30, 50] years
Age>50 and Age<=65	(50, 65] years
Age>65	Above 65 years

8. **(3 points)** For congress 90 and the states of “CA”, “NY”, “TX”, “PA”, and “IL”, show the distribution of the number of congress representatives in each of the age groups (from the previous number).



9. **(3 points)** Create a duplicate of the following graph that shows the distribution of age for each of the democrats and republicans in the congress (all years).



## Part 2: Storytelling with Data

10. **(3 points)** Read Chapter 1 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.