Analysis of most liked and disliked programming languages in US states

**Team members: Bradley Cable, Nandita Ramakrishnan, Sumit Tokkar**

The article focuses on identifying the most liked and disliked languages in each states in USA. The data for analysis is obtained from Stack Exchange. Stack Exchange is a question-and-answer Web site for computer programming question. The primary purpose of the Stack Exchange site is to enable users to post questions, answer them and also allows them to vote on both answers and questions.

#required libraries  
library(choroplethr)  
library(choroplethrMaps)  
library(dplyr)  
library(plyr)  
library(openintro)

# Obtaining the data

The data is available is in csv format and consists of following columns - state, language name, count of likes, dislikes and total count. Following displays the structure of the data using str() function.

#read the dataset  
languageDf <- read.csv("like\_dislike\_by\_state.csv")  
  
#display the dataset structure  
str(languageDf)

## 'data.frame': 19962 obs. of 5 variables:  
## $ tag : Factor w/ 721 levels "3d","access-vba",..: 478 296 339 89 392 307 454 293 140 233 ...  
## $ state : Factor w/ 50 levels "AK","AL","AR",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ likes : int 15 14 13 11 12 11 10 6 7 7 ...  
## $ dislikes: int 1 0 1 2 0 0 1 3 0 0 ...  
## $ total : int 16 14 14 13 12 11 11 9 7 7 ...

# Cleaning the data

The dataset contains like and dislike count of each programming language for each of the 50 US states. We are only interested in the most liked and most disliked language in those states. To obtain the most liked language in each state, we will first group the data by state, sort them by descending order of like and ascending of dislike and then get the first row in each of the group. When there is a tie in like count, we obtain the language, which has the least dislike among them.

#get the most liked language in each state  
  
languagelikeByState <- languageDf %>% group\_by(state) %>%   
 arrange(desc(likes),dislikes) %>%  
 aggregate(by = list(languageDf$state), head, 1) %>%   
 mutate(stateName = tolower(abbr2state(state))) %>%  
 select(stateName, tag) %>%   
 rename(region=stateName, value=tag)   
  
languagelikeByState$value <- factor(languagelikeByState$value)

To obtain the most disliked language in each state, we will first group the data by state, sort them by descending order of dislike and ascending of like and then get the first row in each of the group. When there is a tie in dislike count we obtain the language, which has the least like among them.

#get the most liked language in each state  
   
languagedislikeByState <- languageDf %>% group\_by(state) %>%   
 arrange(desc(dislikes),likes) %>%  
 aggregate(by = list(languageDf$state), head, 1) %>%   
 mutate(stateName = tolower(abbr2state(state))) %>%   
 select(stateName, tag) %>%   
 rename(region=stateName, value=tag)   
  
languagedislikeByState$value <- factor(languagedislikeByState$value)

# Result

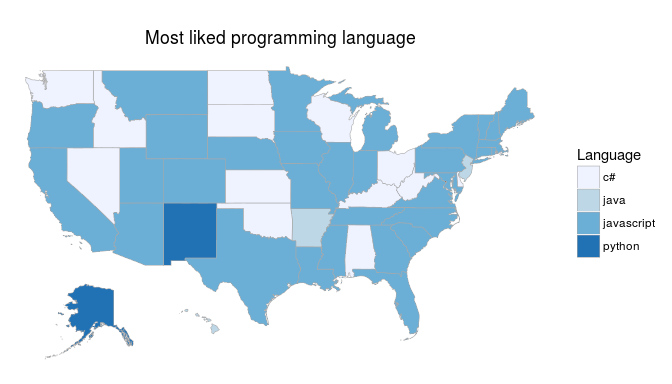
Using the filtered data, we will plot the choropleth map for the most liked and disliked programming language in each state.

### Most liked Programming language

The most liked programming languages are Java, JavaScript, C# and Python. Of the 4 languages, JavaScript is the language that is most liked language in 32 states. Next to JavaScript, C# is the most liked in 13 states. Java is most liked in Arkansas, New Jersey and Hawaii and Python is most liked only in Alaska and New Mexico.

JavaScript has been and is still the most liked programming language in majority of the states in United States.

#plot the map for most liked language  
  
likeMap = StateChoropleth$new(languagelikeByState)  
likeMap$title = "Most liked programming language"  
likeMap$legend = "Language"  
likeMap$show\_labels = FALSE  
  
#display the map  
likeMap$render()

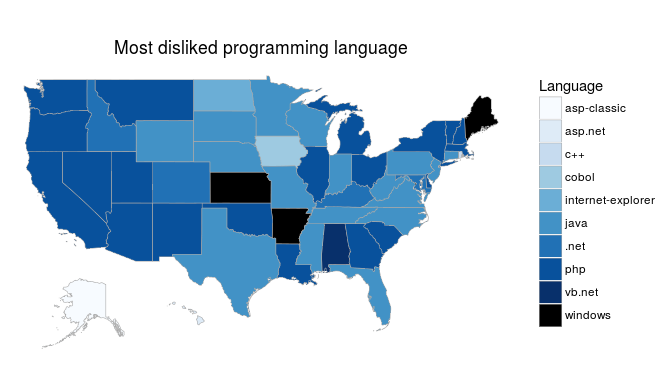
 Figure1: is a choropleth map of the most liked programming language in each state

### 

### Most disliked Programming language

Looking at the map, we can find that there are 10 most disliked languages. asp-classic, asp.net, C++, Cobol, vb.net and internet explorer are most disliked language in anyone of the 50 states. Windows is disliked in 3 states followed by .net in 4 states, java in 17 states. PHP is the most disliked language voted in 20 states with majority of Western states.

#plot the map for most disliked language  
  
dislikeMap = StateChoropleth$new(languagedislikeByState)  
dislikeMap$title = "Most disliked programming language"  
dislikeMap$legend = "Language"  
dislikeMap$show\_labels = FALSE  
  
#display the map  
dislikeMap$render()

 Figure2: is a choropleth map of the most disliked programming language in each state

Looking at both the most liked and disliked language data, we can conclude that JavaScript is most liked and php the most disliked language in majority of states. Java being a popular and widely used language is found in both liked and disliked language list with likes in just 2 states and dislikes in 17 states.