Exam 1 Practice Problems

Christopher Hainzl

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(dslabs)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggpubr)
```

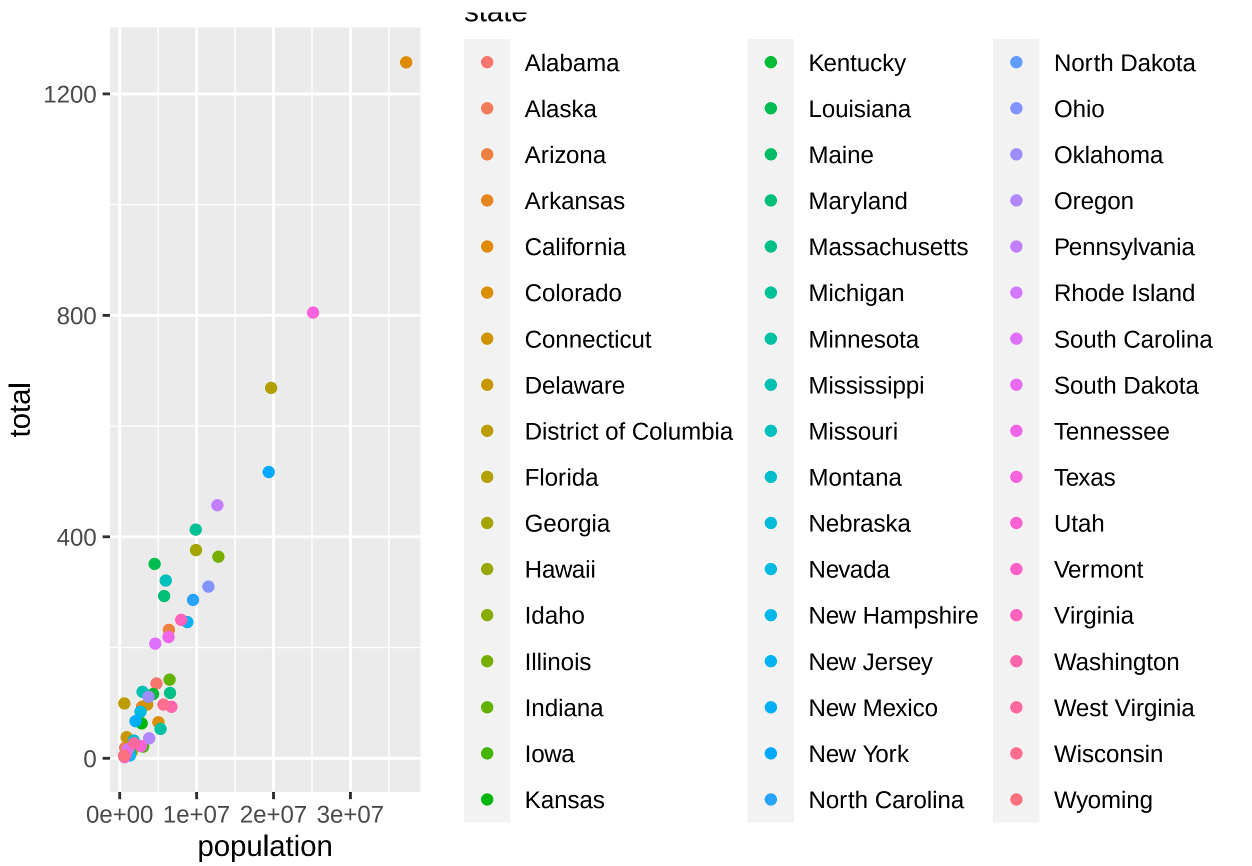
```
## Loading required package: ggplot2
```

Several of your friends live in Europe and are offered jobs at a US company with many locations all across the country. The job offers are great but news with headlines such as America is one of 6 countries that make up more than half of guns deaths worldwide have them worried. You want to convince your friends that the US is a large and diverse country with 50 very different states as well as the District of Columbia (DC). You want to recommend some states for each friend knowing that some like hiking, while others would like to be close to several large cosmopolitan cities. Use data from the US murders data set: 1. What is the state with the most murders? Would you say this is the most dangerous state? Hint: Make a plot showing the relationship between population size and number of murders.

```
murders %>% arrange(desc(total))
```

```
##                   state abb        region population total
## 1            California  CA          West   37253956  1257
## 2                 Texas  TX         South   25145561   805
## 3               Florida  FL         South   19687653   669
## 4              New York  NY     Northeast   19378102   517
## 5          Pennsylvania  PA     Northeast   12702379   457
## 6              Michigan  MI North Central    9883640   413
## 7               Georgia  GA         South    9920000   376
## 8              Illinois  IL North Central   12830632   364
## 9             Louisiana  LA         South    4533372   351
## 10             Missouri  MO North Central    5988927   321
## 11                 Ohio  OH North Central   11536504   310
## 12             Maryland  MD         South    5773552   293
## 13       North Carolina  NC         South    9535483   286
## 14             Virginia  VA         South    8001024   250
## 15           New Jersey  NJ     Northeast    8791894   246
## 16              Arizona  AZ          West    6392017   232
## 17            Tennessee  TN         South    6346105   219
## 18       South Carolina  SC         South    4625364   207
## 19              Indiana  IN North Central    6483802   142
## 20              Alabama  AL         South    4779736   135
## 21          Mississippi  MS         South    2967297   120
## 22        Massachusetts  MA     Northeast    6547629   118
## 23             Kentucky  KY         South    4339367   116
## 24             Oklahoma  OK         South    3751351   111
## 25 District of Columbia  DC         South     601723    99
## 26          Connecticut  CT     Northeast    3574097    97
## 27            Wisconsin  WI North Central    5686986    97
## 28             Arkansas  AR         South    2915918    93
## 29           Washington  WA          West    6724540    93
## 30               Nevada  NV          West    2700551    84
## 31           New Mexico  NM          West    2059179    67
## 32             Colorado  CO          West    5029196    65
## 33               Kansas  KS North Central    2853118    63
## 34            Minnesota  MN North Central    5303925    53
## 35             Delaware  DE         South     897934    38
## 36               Oregon  OR          West    3831074    36
## 37             Nebraska  NE North Central    1826341    32
## 38        West Virginia  WV         South    1852994    27
## 39                 Utah  UT          West    2763885    22
## 40                 Iowa  IA North Central    3046355    21
## 41               Alaska  AK          West     710231    19
## 42         Rhode Island  RI     Northeast    1052567    16
## 43                Idaho  ID          West    1567582    12
## 44              Montana  MT          West     989415    12
## 45                Maine  ME     Northeast    1328361    11
## 46         South Dakota  SD North Central     814180     8
## 47               Hawaii  HI          West    1360301     7
## 48        New Hampshire  NH     Northeast    1316470     5
## 49              Wyoming  WY          West     563626     5
## 50         North Dakota  ND North Central     672591     4
## 51              Vermont  VT     Northeast     625741     2
```

```
murders %>% group_by(total) %>% ggplot(aes(x = population, y = total, col = state)) + geom_point()
```



Solution: California is the state with the most murders. However, this does not necessarily make California the most dangerous state. The following plot shows that the number of murders is highly correlated with the population of any given state. This type of pattern is to be expected as our population size increases. California, the state with the highest population, also has the highest total number of murders.

2. Add a column to the murder data table called murder_rate with each state's murder rate per 100,000 people. Solution:

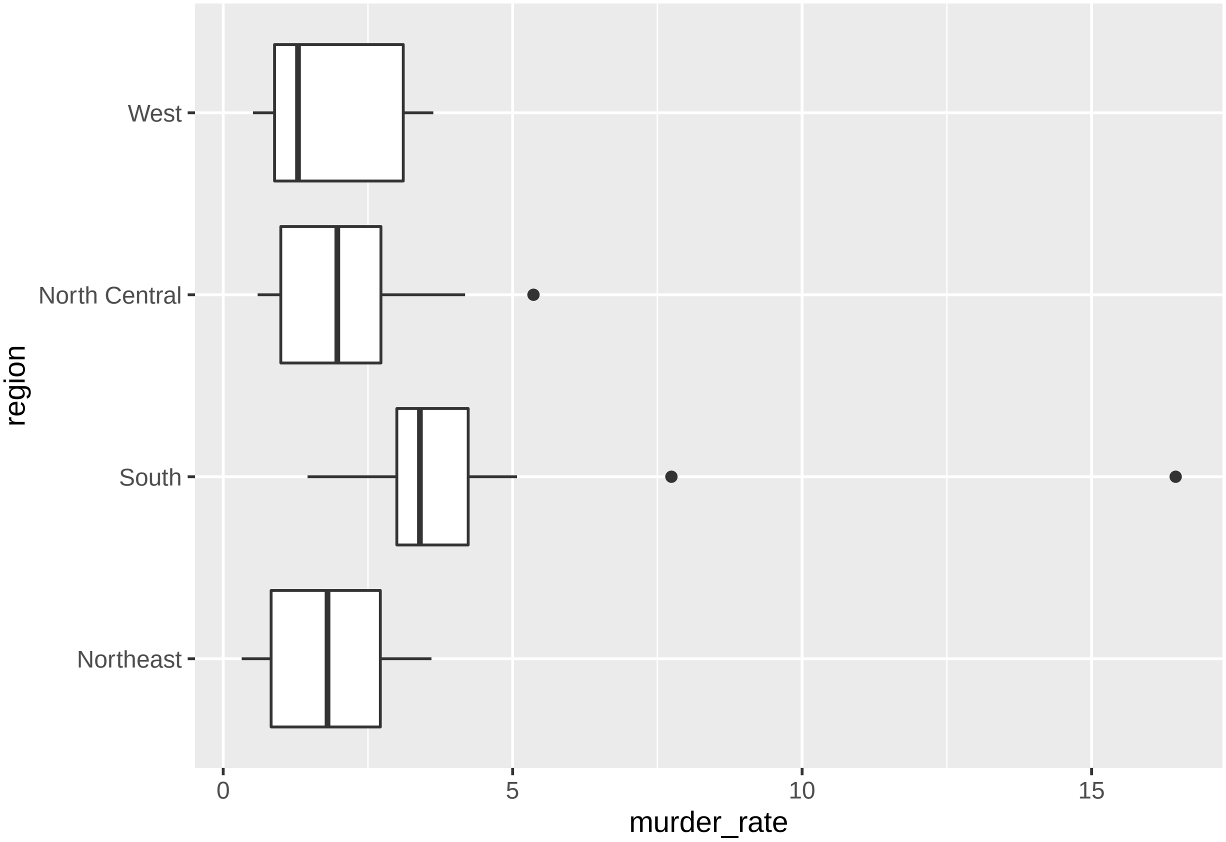
```
rates <- murders %>% mutate(murder_rate = total / (population / 100000))
top_rate <- rates %>% arrange(desc(murder_rate))
top_rate
```

```
##                    state abb        region population total murder_rate
## 1   District of Columbia  DC         South     601723    99  16.4527532
## 2              Louisiana  LA         South    4533372   351   7.7425810
## 3               Missouri  MO North Central    5988927   321   5.3598917
## 4               Maryland  MD         South    5773552   293   5.0748655
## 5         South Carolina  SC         South    4625364   207   4.4753235
## 6               Delaware  DE         South     897934    38   4.2319369
## 7               Michigan  MI North Central    9883640   413   4.1786225
## 8            Mississippi  MS         South    2967297   120   4.0440846
## 9                Georgia  GA         South    9920000   376   3.7903226
## 10               Arizona  AZ          West    6392017   232   3.6295273
## 11          Pennsylvania  PA     Northeast   12702379   457   3.5977513
## 12             Tennessee  TN         South    6346105   219   3.4509357
## 13               Florida  FL         South   19687653   669   3.3980688
## 14            California  CA          West   37253956  1257   3.3741383
## 15            New Mexico  NM          West    2059179    67   3.2537239
## 16          Texas TX         South 25145561 805 3.2013603
## 17       Arkansas AR         South  2915918  93 3.1893901
## 18       Virginia VA         South  8001024 250 3.1246001
## 19         Nevada NV          West  2700551  84 3.1104763
## 20 North Carolina NC         South  9535483 286 2.9993237
## 21       Oklahoma OK         South  3751351 111 2.9589340
## 22       Illinois IL North Central 12830632 364 2.8369608
## 23        Alabama AL         South  4779736 135 2.8244238
## 24     New Jersey NJ     Northeast  8791894 246 2.7980319
## 25    Connecticut CT     Northeast  3574097  97 2.7139722
## 26           Ohio OH North Central 11536504 310 2.6871225
## 27         Alaska AK          West   710231  19 2.6751860
## 28       Kentucky KY         South  4339367 116 2.6732010
## 29       New York NY     Northeast 19378102 517 2.6679599
## 30         Kansas KS North Central  2853118  63 2.2081106
## 31        Indiana IN North Central  6483802 142 2.1900730
## 32  Massachusetts MA     Northeast  6547629 118 1.8021791
## 33       Nebraska NE North Central  1826341  32 1.7521372
## 34      Wisconsin WI North Central  5686986  97 1.7056487
## 35   Rhode Island RI     Northeast  1052567  16 1.5200933
## 36  West Virginia WV         South  1852994  27 1.4571013
## 37     Washington WA          West  6724540  93 1.3829942
## 38       Colorado CO          West  5029196  65 1.2924531
## 39        Montana MT          West   989415  12 1.2128379
## 40      Minnesota MN North Central  5303925  53 0.9992600
## 41   South Dakota SD North Central   814180   8 0.9825837
## 42         Oregon OR          West  3831074  36 0.9396843
## 43        Wyoming WY          West   563626   5 0.8871131
## 44          Maine ME     Northeast  1328361  11 0.8280881
## 45           Utah UT          West  2763885  22 0.7959810
## 46          Idaho ID          West  1567582  12 0.7655102
## 47           Iowa IA North Central  3046355  21 0.6893484
## 48   North Dakota ND North Central   672591   4 0.5947151
## 49         Hawaii HI          West  1360301   7 0.5145920
## 50  New Hampshire NH     Northeast  1316470   5 0.3798036
## 51        Vermont VT     Northeast   625741   2 0.3196211
```

The state with the largest murder rate is District of Columbia. However, this is due to how the District of Columbia is not exactly a state, but rather a city. Individual cities (ex. NYC) can have a larger murder rate compared to the overall one in their respective state.

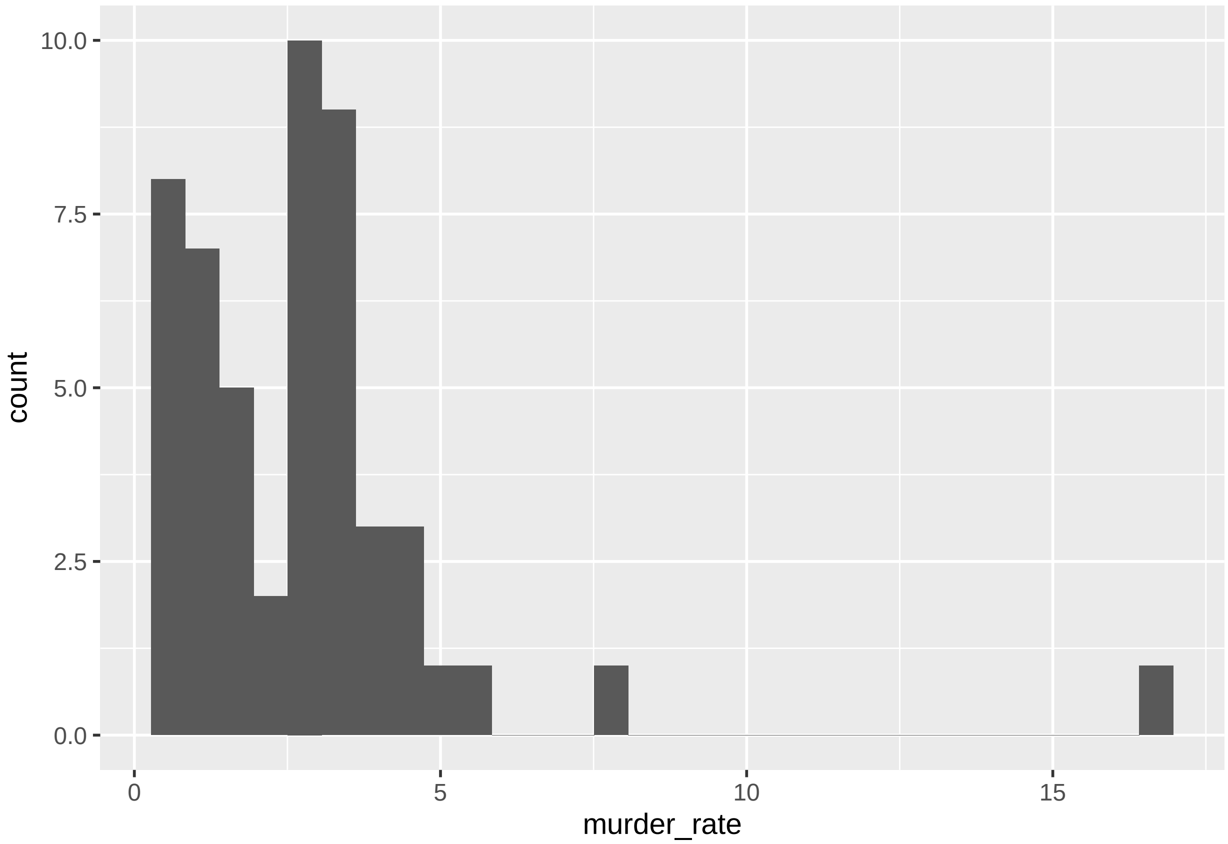
3. Describe the distribution of murder rates across states. How similar are states? How much do murder rates vary by geographical regions?

```r
top_rate %>% ggplot(aes(x = murder_rate, y = region)) +
  geom_boxplot()
```



```
top_rate %>%
  ggplot(aes(x = murder_rate)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Solution: Among all states, the distribution of murder rates is similar in the West, North Central, and Northeast regions. However, the South region has a higher murder rate on average compared to all the other regions.

4. Write a report for your friends reminding them that the US is a large and diverse country with 50 very different states as well as the District of Columbia (DC). Suppose one of your friends loves hiking, one wants to live in a warm climate, and another would like to be close to several large cosmopolitan cities. Recommend a desirable state for each friend. Answers should be a minimum of 1 paragraph and a maximum of 3 paragraphs.

Solution: To my friend who wants to live in a warm climate, I would recommend Hawaii since it is very tropical and has the third-lowest murder rate. To my friend who loves hiking, I would recommend Vermont because not only does it have plenty of hiking trails, but it has the lowest murder rate among all the states in the country. To my friend who wants to be close to several large cosmopolitan cities, I would recommend Massachusetts since it is home to Boston and Cambridge, and also because it has a lower population compared to another state with several cosmopolitan cities like New York (meaning Massachusetts has a lower murder rate than New York).

Question 2 a. Create a function sum_n that for any given value, say n, computes the sum of the integers from 1 to n (inclusive). Use the function to determine the sum of integers from 1 to 200. 1 Solution:

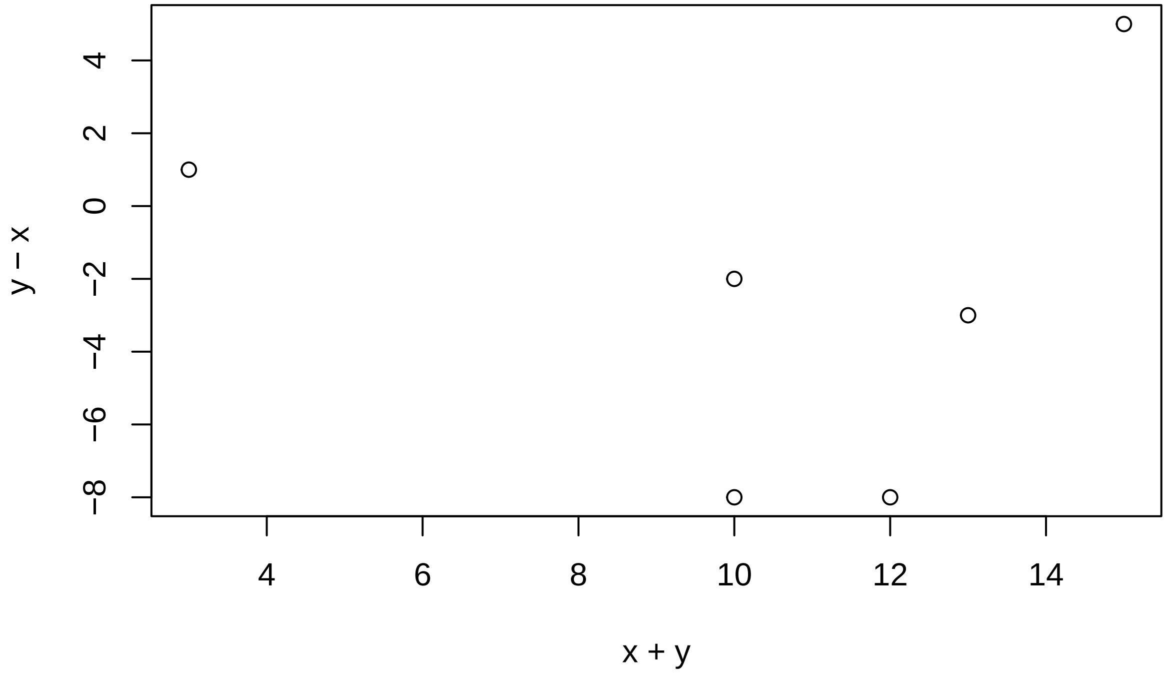
```r
sum_n <- function(n){
  sum <- 0
  for (i in 1:n){
    sum <- sum + i
  }
  sum
}
sum_200 <- sum_n(200)
```

```
print(sum_200)
```

```
## [1] 20100
```

b. Create a function altman_plot that takes two arguments x and y and plots the difference against the sum. Solution:

```
altman_plot <- function(x, y){
  plot(x + y, y - x)
}
d1 <- c(1, 5, 6, 8, 9, 10)
d2 <- c(2, 10, 4, 5, 1, 2)
altman_plot(d1, d2)
```



Things to Consider for exams 1. Major topics included (not an exhaustive list) • Data, and types of variables • Loading data into R • Data Wrangling • Exploratory Data Analysis: Basic Plots and Plotting with ggplot2 • Functions and Iteration 2. Major functions to review (not an exhaustive list) • read.csv() • head() • top_n() • group_by() • summarize() • filter() • mutate() • arrange() • %>% • function(){} • ggplot(); refer to cheatsheet • seq() • c() • data.frame() • which() • which.max() • which.min() • is.na() • print() • for() • ifelse() • mean() • sd() • sqrt() • sum() • Reviewing homework assignments, reading the lecture files, and coding the examples on your own are the best ways to prepare.