STAT 133 Midterm Solutions

1. Consider the following block of code:

```
set.seed(9871234)
x = runif(3)
y = rnorm(2)

set.seed(9871234)
z = rnorm(2)

identical(y, z)
```

What is the return value from the call to identical()? FALSE

Explain. The call to runif between the call to set.seed and rnorm breaks the algorithm for generating random numbers will be in a different positions the two rnorm calls.

2. Short Answer

(a) Write the number 17 in binary

16 8 4 2 1
2^4 2^3 2^2 2^1 2^0
1 0 0 0 1

(b) What is unicode?

a universal standard for representing characters by unique binary digits

(c) Circle the following files that are text files.

(a txt file;) an rda file; a doc file; an xls file; (an HTML file)
binary binary binary

(d) Consider the following function.

```
MAD = function(x, na.rm = FALSE) {
  if (na.rm) {
    x = x[!is.na(x)]
  }
  return( median(abs(x - median(x))))
}
```

Provide the return results for the function call

```
z
[1]  1  2  3 NA

MAD(z)
```

NA because na.rm is FALSE

3. Consider the data frame with Kleiber's animal results, i.e.,

```
> head(animals)

        name      mass       mr
1      Mouse    0.0210     3.60
2        Rat    0.2820    28.10
3  GuineaPig    0.4100    35.10
4     Rabbit    3.3660   159.50
5        Cat    3.0000   152.00
6    Mecaque    4.2000   207.00

> dim(animals)
[1] 19  3
```

(a) Write an R expression to return the subset of `animals` with a `mr` value that is over 1000.

```
animals[ animals$mr > 1000, ]
```

(b) Write an R expression that returns the name of the animal with the largest mass.

```
animals$name[ which( animals$mass == max(animals$mass
```

(c) Suppose there is a second data frame from a more recent experiment that measured the mass and metabolic rates of 23 insects. This data frame is called `insects`. Write an R expression to combine these two data frames into one, called `bothRes`.

```
bothRes = rbind(animals, insects)
```

(d) Write an R expression to add a new column to `bothRes` that indicates whether the observation is an "animal" or "insect". This new column is called `type`

```
bothRes$type = rep( c("animal", insect),
                    c(nrow(animal), nrow(insect)))
```

4. Consider the following function called myFunc().
Also, suppose `aList` contains

```
$x
  x y z     w  v
1 4 d 1  TRUE  6
2 3 j 0 FALSE  2
3 2 l 1 FALSE 17
4 1 b 0  TRUE -3

$y
[1] 3 2 2 2 3

$z
function (x)
{ mean(abs(x)) }
```

Then, for the function call
`myFunc(aList)`
mark the sequence of steps / computations that R carries out in the table below.

myFunc = function(x) {	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n = length(x)	X														
y = vector("list", length = n)		X													
for (j in 1:n) {			X			X			X						
if (class(x[[j]]) == "data.frame") {				X			X			X					
y[[j]] = dim(x[[j]])					X										
} else {															
y[[j]] = NULL								X			X				
}															
}															
return(y)												X			
}															

Rewrite the above code so that it does not use a `for` loop.

```
myFunc = function(x) {
    n = length(x)
    y = vector("list", length = n)
    lapply(x, function(el) {
                if (class(el) = "data.frame") {
                        dim(el)
                } else {
                        NULL
                }
        }
}
```

5. Consider the plot below of bank capital to assets ratio for several countries from 2010 to 2013.

(a) Provide the variables that are included in the plot and their types.

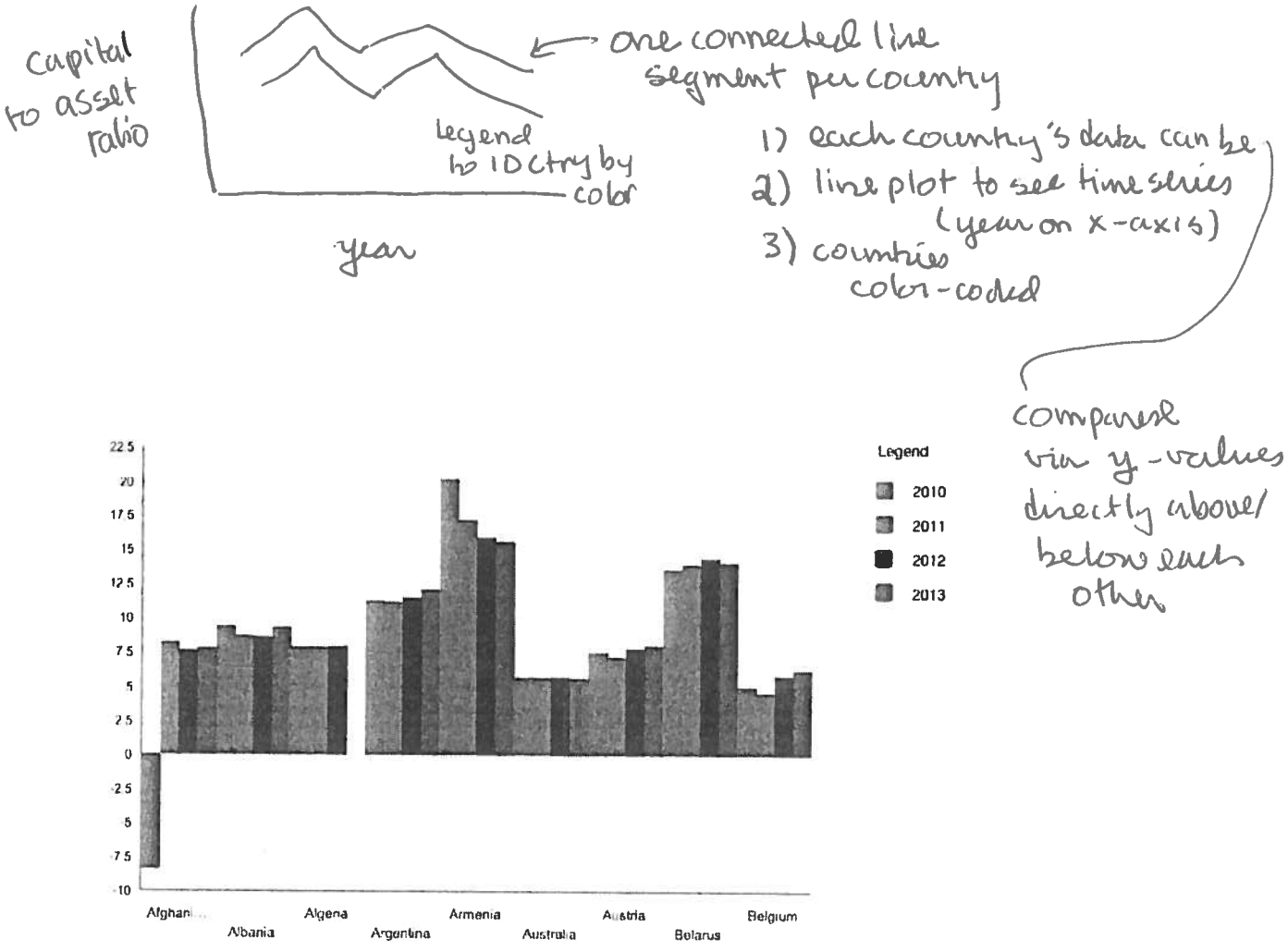
Country - categorical; year - quantitative; bank ratio - quantitative

(b) What is the important comparison to be made with these data?

Compare countries

(c) Below, sketch a plot (not a barplot) that is more appropriate for facilitating this comparison.

(d) Point out the three most important changes you made in your new plot to improve its message.



6. Write a function to convert time to minutes for runners in a race. This function is to be called convertTime(), and it has 3 parameters: `hour`, `min`, and `sec`. These inputs are expected to be vectors where the i^{th} element in each vector corresponds to the hours, minutes, and seconds that the i^{th} runner took to complete the race, respectively. The argument `sec` is optional with a default value of NULL. If `sec` is NULL, then the seconds are not used in the computation of running time. The function returns a vector consisting of the total run time in minutes for each runner. In other words, if a runner took 2 hours, 10 minutes, and 30 seconds to complete the race, then her total run time would be 130.5 minutes. Lastly, if the vectors are not the same length, then the function issues a message and does not perform the computation.

```
convertTime = function(hour, min, sec) {
    if (is.null(sec)) sec = rep(0, length(hour))
    if ( length(hour) != length(min) |
         length(min) != length(sec) ) {
        stop("Mismath of vector length")
    }
    return( hour*60 + min + sec/60 )
}
```

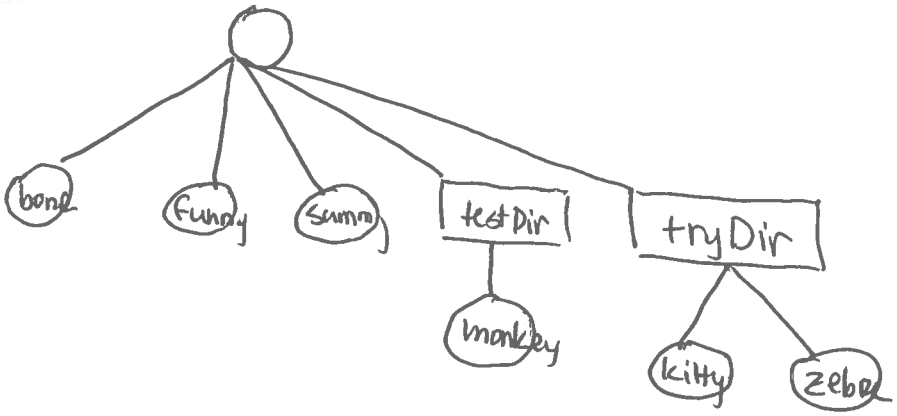
7. Use the following information from the UNIX commands `pwd`, `cd` and `ls` to draw a diagram of the file system with `/Users/nolan/test133/` as the root.

```
$ pwd
/Users/nolan/test133/

$ ls
bone    funny   sammy   testDir/ tryDir/

$ ls testDir/
monkey

$ ls tryDir/
kitty     zebra
```



Next, provide the information that would be printed to the console after each of these UNIX commands

(a) `cd tryDir`
`ls`

kitty zebra

(b) `cp ../testDir/monkey  puppy`
`ls`

kitty puppy zebra

(c) `cd ~/test133/testDir/`
`ls`

monkey