ISCG 8025 Introduction to Data Science Lab Sessions 1 & 2 Due: 21 August 2017

4.

add <https://github.com/kiaorahao/datascienceunitec> to HelloWorld.md file

$ git add HelloWOrld.md

$ git commit -m "submit link"

$ git push origin master

5. <https://github.com/kiaorahao/datascienceunitec/blob/master/HelloWorld.md>

6. a, d, e, h, i

7. c

> x<-c(4, "a", TRUE)

> class(x)

[1] "character"

8.a, d

> x <- c(1,3, 5)

> y <- c(3, 2, 10)

> cbind(x,y)

x y

[1,] 1 3

[2,] 3 2

[3,] 5 10

9. d, e

> x[[2]]

[1] "a"

> x[2]

[[1]]

[1] "a"

10. f

> x+y

[1] 3 5 5 7

> class(x+y)

[1] "integer"

11. b, i, j

a)

> integer(0)<6

logical(0)

b) x[x %in% 1:5] <- 0

[1] 0 0 0 10 12 6

c) x[x>0]<-6

[1] 6 6 6 6 6 6

d) x[x>6]<-0

[1] 3 5 1 0 0 6

e) x[x<6]==0

[1] FALSE FALSE FALSE

f) x[x!=6]<-0

[1] 0 0 0 0 0 6

g) x[x==6]<-0

[1] 3 5 1 10 12 0

h) x[x>=6]<-0

[1] 3 5 1 0 0 0

i) x[x<6]<-0

[1] 0 0 0 10 12 6

j) x[x<=5]<-0

[1] 0 0 0 10 12 6

k) x[x==0]<-6

[1] 3 5 1 10 12 6

12.

> hw1\_data

# A tibble: 153 x 6

Ozone Solar.R Wind Temp Month Day

<int> <int> <dbl> <int> <int> <int>

1 41 190 7.4 67 5 1

2 36 118 8.0 72 5 2

3 12 149 12.6 74 5 3

4 18 313 11.5 62 5 4

5 NA NA 14.3 56 5 5

6 28 NA 14.9 66 5 6

7 23 299 8.6 65 5 7

8 19 99 13.8 59 5 8

9 8 19 20.1 61 5 9

10 NA 194 8.6 69 5 10

# ... with 143 more rows

> hw1\_data[1:2,]

# A tibble: 2 x 6

Ozone Solar.R Wind Temp Month Day

<int> <int> <dbl> <int> <int> <int>

1 41 190 7.4 67 5 1

2 36 118 8.0 72 5 2

13.c

***# Method 1***

row(hw1\_data)

[1] 153

***# Method 2***

nrow(hw1\_data)

[1] 153

14.d

***# Method 1***

> hw1\_data[47,1]

# A tibble: 1 x 1

Ozone

<int>

1 21

***# Method 2***

> hw1\_data[47,'Ozone']

# A tibble: 1 x 1

Ozone

<int>

1. 21

15. a

***# Method 1***

> summary(hw1\_data)

Ozone Solar.R Wind Temp Month Day

Min. : 1.00 Min. : 7.0 Min. : 1.700 Min. :56.00 Min. :5.000 Min. : 1.0

1st Qu.: 18.00 1st Qu.:115.8 1st Qu.: 7.400 1st Qu.:72.00 1st Qu.:6.000 1st Qu.: 8.0

Median : 31.50 Median :205.0 Median : 9.700 Median :79.00 Median :7.000 Median :16.0

Mean : 42.13 Mean :185.9 Mean : 9.958 Mean :77.88 Mean :6.993 Mean :15.8

3rd Qu.: 63.25 3rd Qu.:258.8 3rd Qu.:11.500 3rd Qu.:85.00 3rd Qu.:8.000 3rd Qu.:23.0

Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000 Max. :31.0

NA's :37 NA's :7

***# Method 2***

> sum(is.na(hw1\_data$Ozone))

[1] 37

16. a

***# Method 1***

> colMeans(hw1\_data,na.rm = TRUE)

Ozone Solar.R Wind Temp Month Day

42.129310 185.931507 9.957516 77.882353 6.993464 15.803922

***# Method 2***

> mean(as.numeric(na.omit(hw1\_data$Ozone)))

[1] 42.12931

17. b

> t<-subset(hw1\_data,hw1\_data$Month==5)

> max(t$Ozone,na.rm = TRUE)

Reference: <http://w432111.cool.blog.163.com/blog/static/328831842015020102854905/>

18. b

19. d

Warning message:

In if (x > 5) { :

the condition has length > 1 and only the first element will be used

correct:

x[x>5]<-0

20. a

21. b

22. a

23. b

24. d

25. b

26. b

27. a

Reference: file:///Users/Harry/Downloads/Quize%202%20week%20(answer).pdf

28. 7

***# Method 1***

m<-subset(iris,iris$Species=='virginica')

round(mean(m$Sepal.Length,na.rm = TRUE))

***# Method 2***

round(mean(iris[which(iris$Species == "virginica"),]$Sepal.Length))

Reference:

https://rpubs.com/ryantillis/R\_Programming\_3

29. e?

Sepal.Length Sepal.Width Petal.Length Petal.Width

5.843333 3.057333 3.758000 1.199333

30. 34