Assignment 1

Ingrid Aukrust

2022-11-02

Exercise 1

a)

```
cov = c(0.25,0.12,0.36,0.16,0.12,0.77,0.66,0.35,0.36,0.66,3.68,0.74,0.16,0.35,0.74
(covmatrix = matrix(cov,nrow=4,ncol=4,byrow=TRUE))
```

```
##      [,1] [,2] [,3] [,4]
## [1,] 0.25 0.12 0.36 0.16
## [2,] 0.12 0.77 0.66 0.35
## [3,] 0.36 0.66 3.68 0.74
## [4,] 0.16 0.35 0.74 1.18
```

```
sum(covmatrix)
```

```
## [1] 10.66
```

The estimated variance of a sum score of $X_1 + X_2 + X_3 + X_4$ is the sum of the variance.

$$Var(X_1 + X_2 + X_3 + X_4) = \sum_{j=1}^{4}\sum_{i=1}^{4} = 10,66$$

b)

The estimated correlation matrix from the correlation matrix is

```
cov2cor(covmatrix)
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.2735054 0.3753259 0.2945839
## [2,] 0.2735054 1.0000000 0.3920800 0.3671822
## [3,] 0.3753259 0.3920800 1.0000000 0.3551132
## [4,] 0.2945839 0.3671822 0.3551132 1.0000000
```

Correlation is calculated from covariance with

$$\rho_X, Y = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

c)

From the correlation matrix we can recognize pairs of items that seeme most related, which might be that they measure the same thing. On the other hand we can see sets of items that seemes to measure different things. In this case we see that the there is no correlation between any of the items that seem strong, while the strongest relationship is a correlation of 0.39 between X_2 and X_3. Even though this is not concidered a strong correlation this might indicate that they to some degree measure the same thing. Still, the rest of the items seem to measure different things, as the correlation is low.

d)

To better have an understanding of the relationships between the variables, …,jh khgk.

Exercise 2

From the table of proportion correctly answered from each item, the calculated estimete of chronbacs alpha is:

$$\frac{7}{6} \times (1 - \frac{0.3 - 0.52}{3,67} + \frac{0.2 - 0.52}{3,67} + \frac{0.94 - 0.52}{3,67} + \frac{(0.69 \times 2) - 0.52}{3,67} + \frac{(0.52 \times 2) - 0.52}{3,67} = 0.74$$

0.52 is the mean of the proportions correctly answered across items.

Exercise 3

a)

(See R code) Items 1 and 2 I would score according to ranking, where I assign integers $1 - 5$ and $1 - 4$ to respectfully “Never” - “Every day” and “I have never heard of this” - “I am familiar with this and I would be able to explain this well”. The third part contains only nominal items, which are categories and therefore not subject to ranking. In this case they are hard to put a number on in terms of ranking, but I will put a number on to place them in categories.

b)

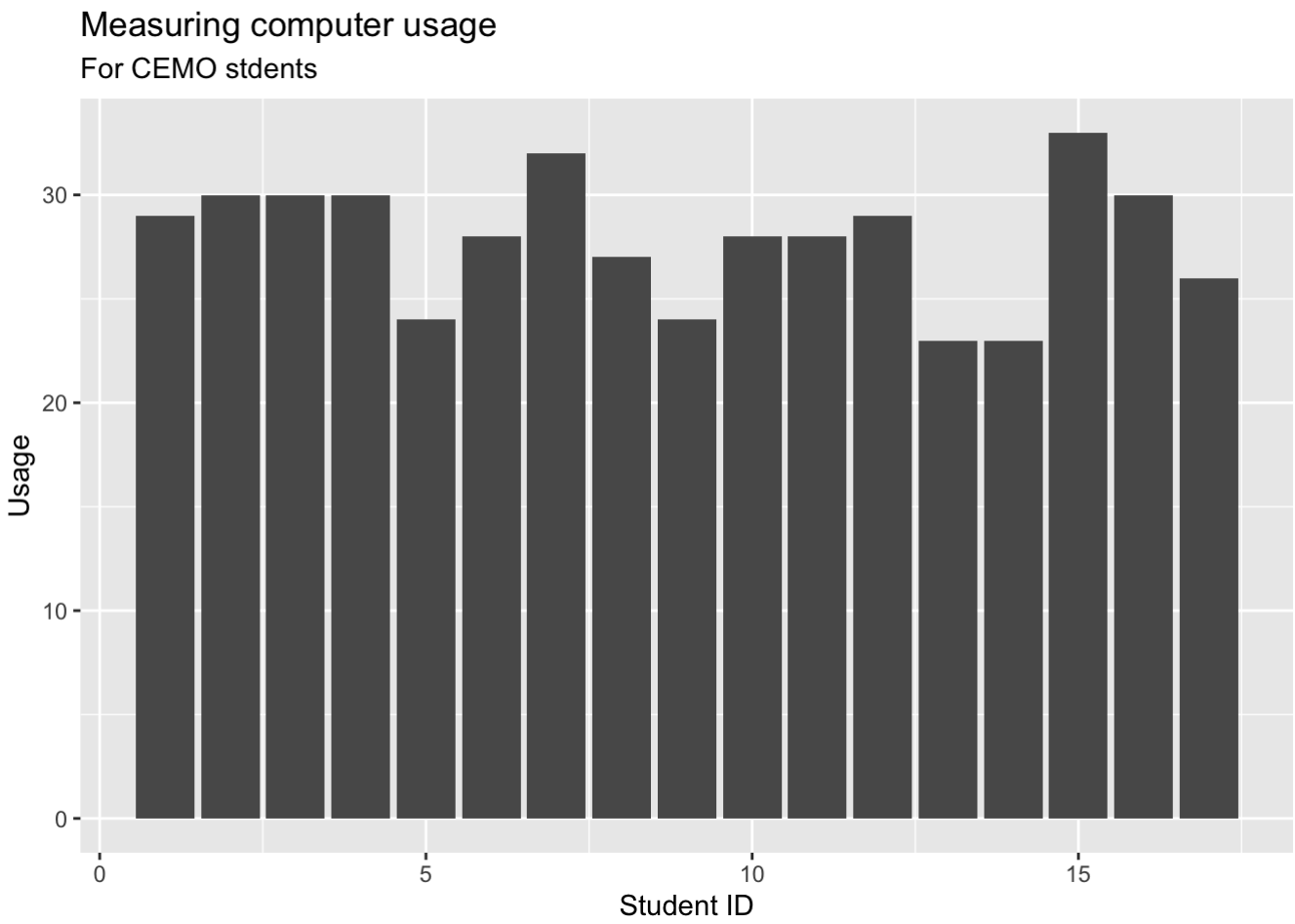
(See R code)

c)

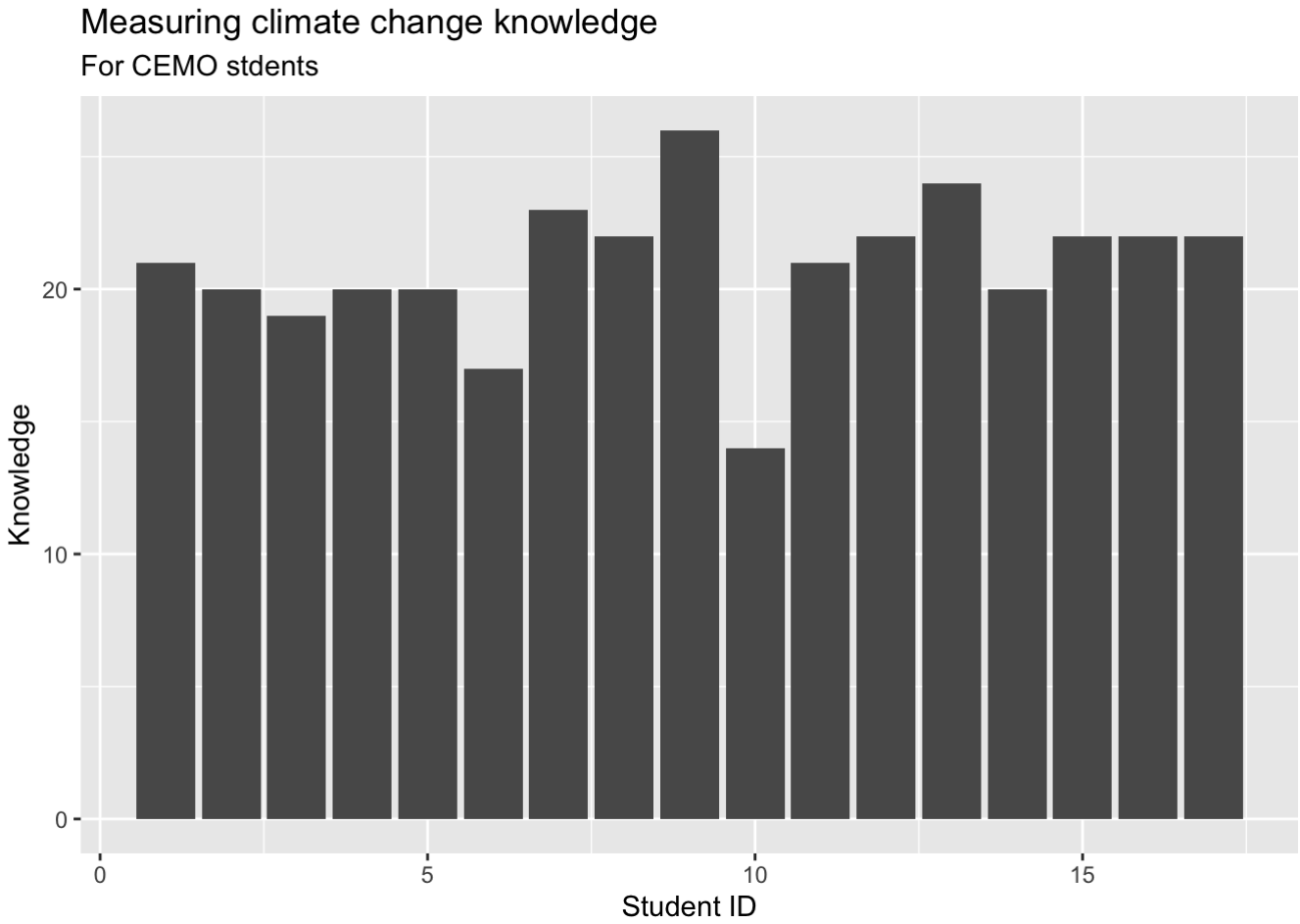
(See R code)

d)

```
(flights_visual <- ggplot(data = y, aes(x = ID, y = computer_sum))+
    geom_col()+
    labs(
      title = "Measuring computer usage",
      subtitle = "For CEMO stdents",
      x = "Student ID",
      y = "Usage"))
```




```
(flights_visual <- ggplot(data = y, aes(x = ID, y = earth_sum))+
    geom_col()+
    labs(
        title = "Measuring climate change knowledge",
        subtitle = "For CEMO stdents",
        x = "Student ID",
        y = "Knowledge"))
```



I chose bars for both visualizations because they easily present the data in an intuitive way, as well as allowing comparing between students.