CSC 424 Homework3 Part 2

2.

Source Code:

```
library(psych)
ds <- read.csv("/Users/Yiyang/Documents/CSC 424/Final Dataset/Dataset.csv", sep = ",", header = TRUE)
dm <- ds[, c(2: 9, 11)]
dm.pr1 <- prcomp(dm, center = TRUE)
summary(dm.pr1)
dm.cf <- principal(dm, rotate = "varimax", nfactors = 2, score = TRUE)
print(dm.cf$loadings, cutoff = 0.4, sort = TRUE)
```

Output:

```
Importance of components:
                          PC1      PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9
Standard deviation     64.9870 11.07572 8.37263 2.40668 1.47570 1.01853 0.80667 0.23692 0.09441
Proportion of Variance  0.9543  0.02772 0.01584 0.00131 0.00049 0.00023 0.00015 0.00001 0.00000
Cumulative Proportion   0.9543  0.98196 0.99780 0.99911 0.99960 0.99984 0.99999 1.00000 1.00000

Loadings:
                               RC1    RC2
mean_temp_total                 0.786
mean_dew_point_f_total          0.921
mean_wind_speed_mph             0.666
mean_humidity                          0.615
mean_visibility_miles                 -0.612
precipitation_inches                   0.631
cloud_cover                            0.761
wind_dir_degrees                      -0.560
mean_sea_level_pressure_inches -0.409

                 RC1   RC2
SS loadings    2.496 2.207
Proportion Var 0.277 0.245
Cumulative Var 0.277 0.523
```

After applying PCA to this dataset, I get 2 principle component that will explain 98% total variation, then I apply CFA to the dataset, I get 2 common factors. Since we will discuss how natural factors effect the duration of biking, from the variables of the first common factor, it could represent the human comfort under the effect of some natural elements; the second common factor could represent the possibility of rain. In my view, these two factors are very suitable for the dataset.