```
#===============================#
# Multivariate Statistik WS14/15
# Exercise 4
# Andreas Blatt
#===============================#

### Read data and load packages
library(StatDA)
base <- read.csv("http://www.statistik.tuwien.ac.at/public/filz/students/multi
/cardata.csv")
base2 <- read.csv("http://www.statistik.tuwien.ac.at/public/filz/students/multi/flu.csv")

# Data First Part: remove NAs and unused columns (1-9, 15, 16 and 18)
base.na <- na.omit(base)
work <- base.na[,-c(1:9,15,16,18)]

# Data Second Part:
work2 <- base2[-c(1:160),-c(26)]
work2 <- as.matrix(log(work2[,2:29]))

### Part 1
## Principal Component Analysis
pca <- prcomp(scale(work), k=2)
summary(pca)
biplot(pca)

# The first component explains 99% of total variance

## Factor analysis
fa <- factanal(scale(work),factors=2,scores="regression")

# Add a biplot method to "factanal"
biplot.factanal <- function(fa.fit,...)
{
  x = fa.fit$scores[,1:2]
  y = fa.fit$loadings[,1:2]
  biplot(x,y,...)
}

biplot(fa)
fa$loadings[,2]/fa$loadings[,1] # slope of the vectors within the (factor 1, factor
2)-plane

# Differences to PCA: vectors not orthogonal and linear combinations, that maximize
shared
# portion of the variance. It focuses more on the covariance and correlation between
variables.

# - loadings:
# - scores:
# - uniqueness: yes
print(fa)

# How some results are computed:
SSQ_loadings <- colSums(fa$loadings^2)
SSQ_loadings
prop_Var <- (SSQ_loadings)/nrow(unclass(fa$loadings))
prop_Var
```

```
## Principal Factor Analysis
pfa <- pfa(scale(work),factors=2,scores="regression")
biplot(pfa)



### Part 2
# flu.trend <- function(i)
# {
#   ts1 <- ts(base2[,i],start=c(2002,52),frequency=52)
#   T1 <- time(ts1)
#   M1 <- factor(cycle(ts1))
#   lm1 <- lm(ts1~T1+M1)
#   p1 <- predict(lm1,newdata=data.frame(T1,M1))
#
#   plot(T1,p1,type="l",col="red",main=colnames(base2)[i])
#   abline(lm1$coefficients[1],lm1$coefficients[2],col="blue")
#   points(T1,base2[,i])
# }

# par(mfrow=c(2,2))
# for (i in sample(2:ncol(base2),4)){flu.trend(i)}
# par(mfrow=c(1,1))


pfa.flu <- pfa(scale(work2),factors=5,scores="regression")
print(pfa.flu)$PVAL
biplot(pfa.flu)

fa.flu <- factanal(scale(work2),factors=21,scores="regression")
print(fa.flu)$PVAL
biplot(pfa.flu)
```