

# Įskaitinių eismo įvykių Lietuvos keliuose analizė

*Kornelijus Samsonas, Linas Šyvis*

Reikalingi užduotims paketai.

```
library(fpp)
```

Nuskaityti duomenis.

```
rawdata <- read.csv2("2003-2015.csv", header = T)
agreguoti <- read.csv2("agreguoti2009-2015.csv", header = T)
gyventojai <- read.csv2("gyventojai.csv", header = T, skip = 1)
kor <- read.csv2("koreliacijos.csv", header = T)
```

Apžvelgiame duomenis

```
sapply(c(rawdata, agreguoti, gyventojai), head)
```

```
##      Vieta Data Laikas Dalyviu_skaicius Zuvusiu_skaicius Suzeistu_skaicius
## [1,]      1  20   1442              1              0              1
## [2,]      1  22   1442              2              0              2
## [3,]      1  26   3113              2              0              2
## [4,]      1  57   5797              2              0              1
## [5,]      1  75    754              2              0              1
## [6,]      1  76   5459              2              0              1
##      Dalyvavusiu_transporto_priemoniu_skaicius
## [1,]
## [2,]
## [3,]
## [4,]
## [5,]
## [6,]
##      Apgadintu_transporto_priemoniu_skaicius Kelio_reiksme Dangos_rusis
## [1,]              1              NA              1
## [2,]              1              NA              1
## [3,]              1              NA              4
## [4,]              NA              1              1
## [5,]              1              3              1
## [6,]              1              1              1
##      Dangos_bukle Paros_metas Meteorologines_salygos Kelio_kreive
## [1,]              1              1              2              4
## [2,]              1              1              1              4
## [3,]              1              1              1              4
## [4,]              3              3              2              1
## [5,]              3              1              2              3
## [6,]              3              3              2              4
##      Atitvarai Eismo_ivykio_dalyvio_kategorija
## [1,]              0              2
## [2,]              0              2
## [3,]              0              2
## [4,]              0             11
## [5,]              0              2
## [6,]              0             11
##      KET_pazeidimas_del_kurio_ivyko_eismo_ivykis Lytis Pilietybe
```

```
## [1,] 22 1 18
## [2,] 22 3 18
## [3,] 22 3 18
## [4,] 35 3 18
## [5,] 22 3 18
## [6,] 35 3 18
##      Gimimo_data Neigalumas Saugos_dirzas Oro_pagalviusuveikimas
## [1,] 6250 0 3 NA
## [2,] 4286 0 NA NA
## [3,] 8136 0 NA NA
## [4,] 2163 0 NA NA
## [5,] 14506 0 NA NA
## [6,] 4206 0 NA NA
##      Vairuotojo_kvalifikacija Stazas Blaivumas Metai Data Vieta
## [1,] 1 12 3 2009 1 28
## [2,] 1 32 3 2009 1 2
## [3,] 1 22 5 2009 1 3
## [4,] NA NA 5 2009 1 4
## [5,] 1 4 3 2009 1 8
## [6,] NA NA 5 2009 1 27
##      Visi_ivykiai Del_neblaiviu_vairuotoju_kaltes Suzeistieji Zuvusieji
## [1,] 262 17 303 28
## [2,] 13 0 12 3
## [3,] 7 0 8 0
## [4,] 1 0 0 1
## [5,] 1 0 2 0
## [6,] 3 0 1 2
##      Vieta Data Visi Vyras Moterys
## [1,] 4 2003 3431497 1601098 1830399
## [2,] 4 2004 3398929 1584018 1814911
## [3,] 4 2005 3355220 1562264 1792956
## [4,] 4 2006 3289835 1528463 1761372
## [5,] 4 2007 3249983 1506794 1743189
## [6,] 4 2008 3212605 1487419 1725186
```

## Tvarkome duomenis

Paliekame tik 5 didžiuosius miestus

```
miestai <- c("Vilniaus m. sav.", "Kauno m. sav.", "Klaipėdos m. sav.", "Siauliu m. sav.", "Panevezio m. sav.")
data <- subset(rawdata, Vieta %in% miestai)
```

Turime stulpelius “Gimimo data” ir “Stazas”. Jų reikšmės tikėtina stipriai koreliuoja, patikrinkime. Pertvarkysime stulpelio “Gimimo data” reikšmes taip, kad matytume eismo įvykio dalyvio amžių ir patikrinsime koreliaciją su “Stazo” reikšmėmis.

Stulpelį “Gimimo data” pertvarkome į “Amzius”.

```
Amzius <- as.numeric(format(Sys.Date(), format="%Y")) - as.numeric(format(as.Date(data$Gimimo_data, format="%Y-%m-%d"), format="%Y"))
data <- data[,-22]
data <- cbind(data, Amzius)
head(data$Amzius)
```

```
## [1] 36 45 73 33 39 40
```

Patikrinsime koreliaciją.

```
sum(is.na(data$Stazas))
```

```
## [1] 8184
```

```
cor(data$Amzius, data$Stazas)
```

```
## [1] NA
```

Reikia pašalinti NA, kad pamatuot cor.

```
table(data$Blaivumas)
```

```
##
## Apsvaiges nuo narkotiniu, psichotropiniu ar kitu psichika veikianciu medziagu
##                                     14
##                               Atsisake buti patikrintas
##                                     12
##                               Blaivus
##                               21615
##                               Neapsvaiges
##                               0
##                               Neblaivus
##                               2218
```

Matome, kad stulpelį “Blaivumas” sudaro 6 skirtingos reikšmės. Kadangi, sutrauksime jas visas į “Blaivus” ir “Neblaivus”.

```
data[data[, "Blaivumas"] %in% c("Apsvaiges nuo narkotiniu, psichotropiniu ar kitu psichika veikianciu m
data$Blaivumas <- factor(data$Blaivumas)
table(data$Blaivumas)
```

```
##
##   Blaivus Neblaivus
##   21615      2244
```

Matome, kad yra trūkstamų reikšmių, reikia jas pašalinti.

```
ndata <- data
data <- data[-which(is.na(data$Blaivumas)),]
"ndata <- ndata[-lapply(lapply(data, is.na), which),]"
```

```
## [1] "ndata <- ndata[-lapply(lapply(data, is.na), which),]"
```

5 didžiausius Lietuvos miestus lyginsime pagal įvykius 1000-iui gyventojų.

Sukuriame rodiklį 1000 gyventojų.

```
tukst <- gyventojai[,3]/1000
gyventojai <- cbind(gyventojai, tukst)
```

## Įvykių sk. 1000-iui gyventojų didžiausiuose miestuose grafikas.

**\*\*Išsaugome didžiųjų miestų ir Lietuvos įvykių skaičių ir gyventojų laiko eilutes.**

```
ivykiusk <- function(i)
{
  ts(agreguoti[agreguoti$Vieta %in% i, "Visi_ivykiai"], start=c(2009,1), frequency = 12)
}
```

```
gyvsk <- function(i)
```

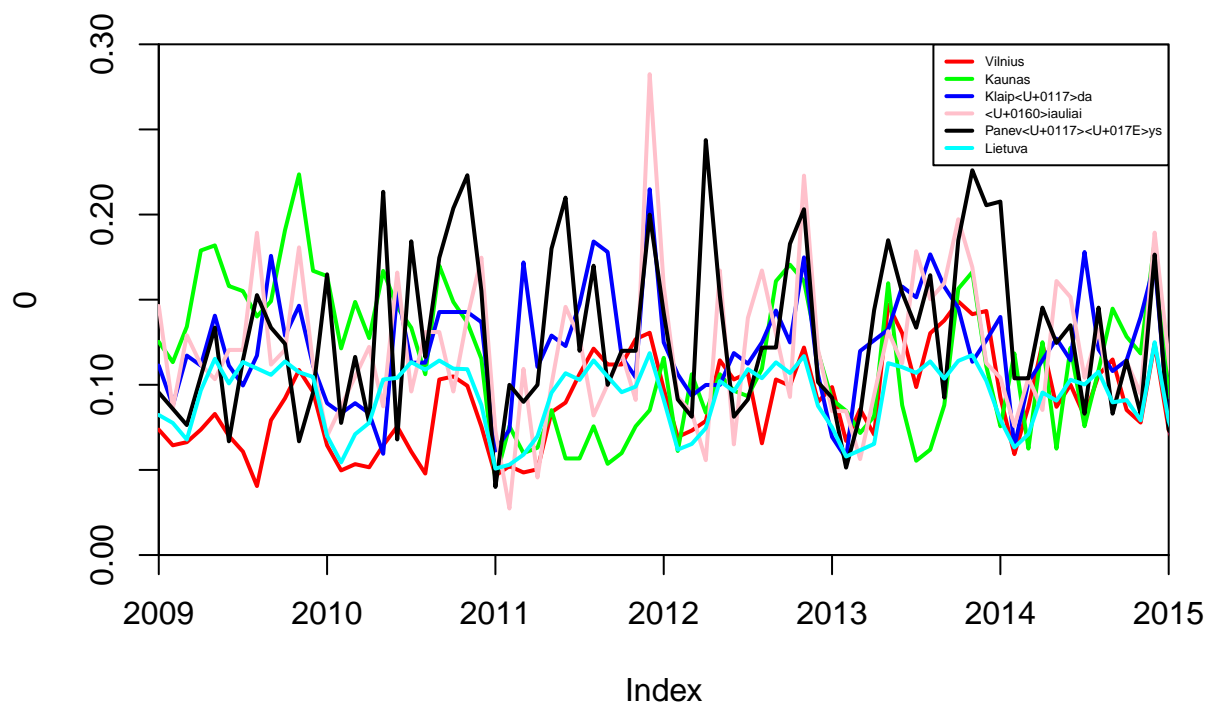
```
{
  ts(rep(gyventojai[gyventojai$Vieta == i & gyventojai$Data %in% 2009:2015, "Visi"], each = 12), start = 2009)
}

ivykiai <- lapply(c(mietai, "Lietuvos Respublika"), ivykisk)
names(ivykiai) <- sprintf(c(mietai, "Lietuvos Respublika"))
gyven <- lapply(c(mietai, "Lietuvos Respublika"), gyvsk)
names(gyven) <- sprintf(c(mietai, "Lietuvos Respublika"))
```

Brėžiame grafiką.

```
plot(0, xlim = c(2009,2015), ylim = c(0,0.3), yaxs="i", xaxs="i", main = "Įvykių skaičius 1000-iui gyventojui",
lines(ivykiai$`Vilniaus m. sav.`*1000/gyven$`Vilniaus m. sav.` , col = "red", lwd = 2)
lines(ivykiai$`Kauno m. sav.`*1000/gyven$`Kauno m. sav.` , col = "green", lwd = 2)
lines(ivykiai$`Klaipėdos m. sav.`*1000/gyven$`Klaipėdos m. sav.` , col = "blue", lwd = 2)
lines(ivykiai$`Siauliu m. sav.`*1000/gyven$`Siauliu m. sav.` , col = "pink", lwd = 2)
lines(ivykiai$`Panevėžio m. sav.`*1000/gyven$`Panevėžio m. sav.` , col = "black", lwd = 2)
lines(ivykiai$`Lietuvos Respublika`*1000/gyven$`Lietuvos Respublika` , col = "cyan", lwd = 2)
legend("topright", col = c("red", "green", "blue", "pink", "black", "cyan"), lty = 1, lwd = 2, legend = TRUE)
```

**<U+012E>vyki<U+0173> skai<U+010D>ius 1000–iui gyventoj<U+0173>**



**Skaičiuojam neblaivius**

```
"neblaivus <- as.numeric(0)
blaivus <- as.numeric(0)
for(i in 1:5)
neblaivus[i] <- nrow(data[data[,25] == Neblaivus & data[,1] == mietai[i],]) / gyventojai[gyventojai[,1] == mietai[i],25]
rbind(mietai, round(neblaivus, digits = 3))
```

```

for(i in 1:5)
neblaivus[i] <- nrow(data[data[,25] == Neblaivus & data[,1] == miestai[i],])
for(i in 1:5)
blaivus[i] <- nrow(data[data[,25] == Blaivus & data[,1] == miestai[i],])

tab <-matrix(c(59, 434, 25, 269, 9, 138, 13, 90, 3, 93), nrow=5, byrow=T)
colnames(tab) <- c(taip, ne)
rownames(tab) <- miestai
prop.test(tab)"

```

```
## [1] "neblaivus <- as.numeric(0)\nblaivus <- as.numeric(0)\nfor(i in 1:5)\nneblaivus[i] <- nrow(data[data[,25] == neblaivus & data[,1] == miestai[i],])\nblaivus[i] <- nrow(data[data[,25] == blaivus & data[,1] == miestai[i],])"
```

Matome, kad  $p\text{-value} < 0.05$ , vadinasi  $H_0$  atmetam. Proporcijos statistiškai reikšmingai skiriasi. Patikrinkime tarp kurių miestų proporcijos yra statistiškai reikšmingai skirtingos.

```
pairwise.prop.test(tab, p.adjust.method = "none")
```

Proporcijos statistiškai reikšmingai skiriasi tarp Šiaulių m. sav. ir Panevėžio m. sav. ir tarp Vilniaus m. sav. ir Panevėžio m. sav. Patikrinkime, ar Šiauliuose/Vilniuje įvykių dėl neblaivių vairuotojų kaltės yra statistiškai reikšmingai daugiau Panevėžyje.

```

tab1 <-matrix(c(13, 90, 3, 93), nrow=2, byrow=T)
prop.test(tab1, alternative = "greater")

```

```

##
## 2-sample test for equality of proportions with continuity
## correction
##
## data:  tab1
## X-squared = 4.8441, df = 1, p-value = 0.01387
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.02366318 1.00000000
## sample estimates:
##   prop 1    prop 2
## 0.1262136 0.0312500

```

```

tab2 <-matrix(c(59, 434, 3, 93), nrow=2, byrow=T)
prop.test(tab2, alternative = "greater")

```

```

##
## 2-sample test for equality of proportions with continuity
## correction
##
## data:  tab2
## X-squared = 5.7651, df = 1, p-value = 0.008174
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.04436963 1.00000000
## sample estimates:
##   prop 1    prop 2
## 0.1196755 0.0312500

```

Paaikėjo, kad Šiauliuose/Vilniuje įvykių dėl neblaivių vairuotojų kaltės yra statistiškai reikšmingai daugiau negu Panevėžyje.

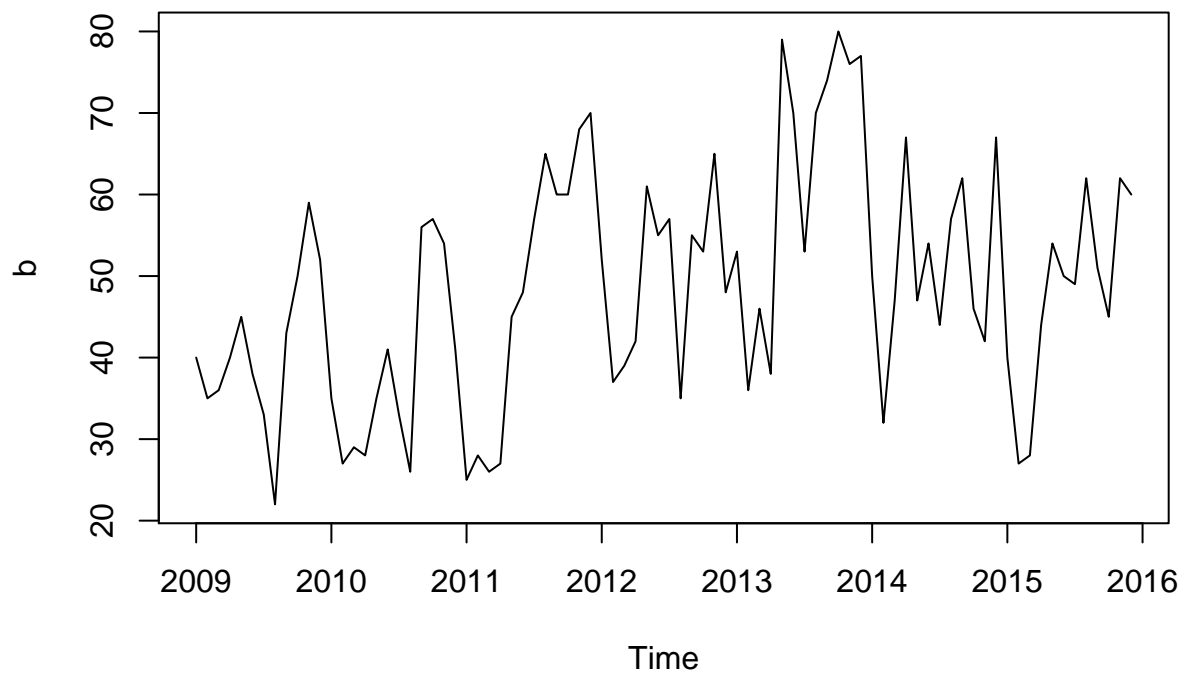
## Tikrinam įvykių priklausomybę nuo amžiaus

```
stazas <- cut(data[,24], breaks = c(0, 10, 20, 30, 40, 50, 60))  
table(stazas)
```

```
## stazas  
## (0,10] (10,20] (20,30] (30,40] (40,50] (50,60]  
## 7364 4022 2155 1036 317 72
```

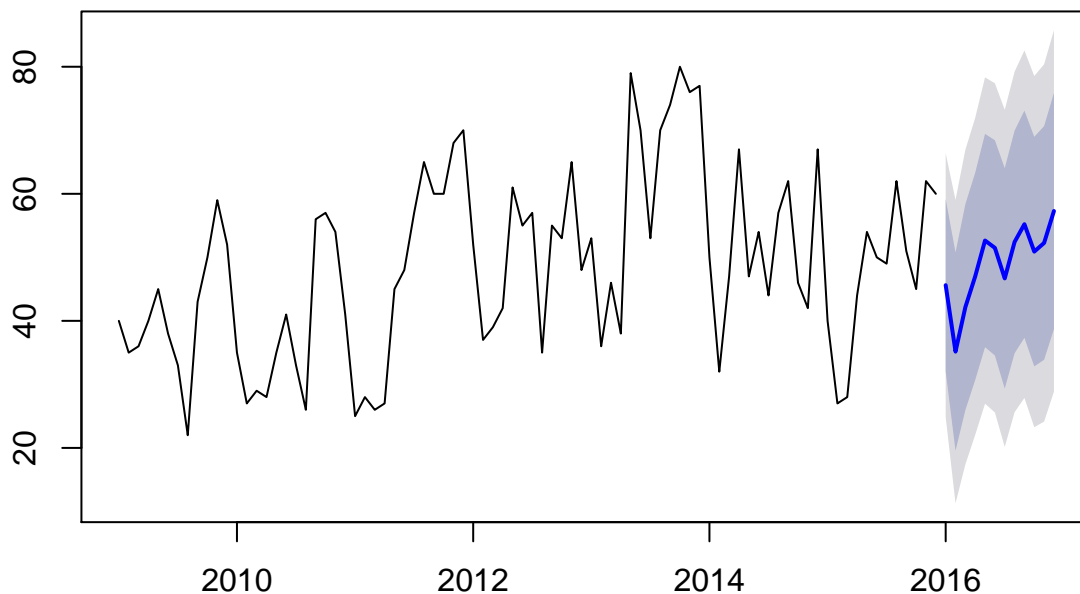
## Forecastinam autoįvykius

```
b <- ts(agreguoti[agreguoti[,3] == "Vilniaus m. sav.", 4], start=c(2009,1), frequency = 12)  
plot(b)
```

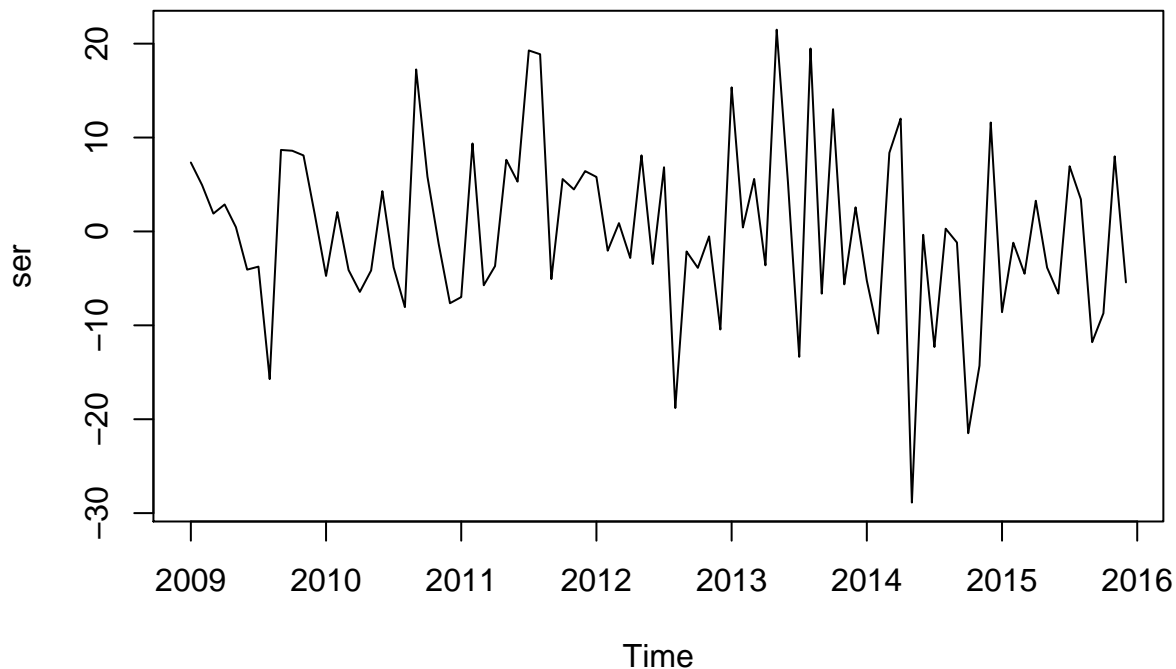


```
z <- auto.arima(b)  
plot(forecast(z, h = 12))
```

## Forecasts from ARIMA(1,1,3)(1,0,2)[12]



```
#salinam sezona
liek <- ts(z$residuals,start=c(2009,1), frequency = 12)
stl <- stl(z$residuals, s.window="periodic")
ser <- liek - stl$time.series[, "seasonal"]
plot(ser)
```



Tikrinsime hipoteze ar Vilniaus mieste eismo ivykiu kaltininkai dazniau yra Vyrai nei moterys.

$H_0 = V \text{ sk.} =$

```
Vilniusm <- rawdata[rawdata[,1] == "Vilniaus m. sav.",]
Kaunasm <- rawdata[rawdata[,1] == "Kauno m. sav.",]
```

```
Siauliaim <- rawdata[rawdata[,1] == "siauliu m. sav.",]
Klaipedam <- rawdata[rawdata[,1] == "Klaipedos m. sav.",]
Panevezysm <- rawdata[rawdata[,1] == "Panevezio m. sav.",]
```

### Kaltininku procentas miestuose (Vyrų/Moterys)

Kaltininku procentas Vilniaus mieste.

```
vyruskV<-sum(na.omit(as.numeric(Vilniusm[,19]=="Vyras")))
moteruskV<-sum(na.omit(as.numeric(Vilniusm[,19]=="Moteris")))
visoV<-(vyruskV+moteruskV)
vyruprocV=vyruskV/visoV*100
moteruprocV=moteruskV/visoV*100
```

Kaltininku procentas Kauno mieste.

```
vyruskK<-sum(na.omit(as.numeric(Kaunasm[,19]=="Vyras")))
moteruskK<-sum(na.omit(as.numeric(Kaunasm[,19]=="Moteris")))
visoK<-(vyruskK+moteruskK)
vyruprocK=vyruskK/visoK*100
moteruprocK=moteruskK/visoK*100
```

Kaltininku procentas Siauliu mieste.

```
vyruskS<-sum(na.omit(as.numeric(Siauliaim[,19]=="Vyras")))
moteruskS<-sum(na.omit(as.numeric(Siauliaim[,19]=="Moteris")))
visoS<-(vyruskS+moteruskS)
vyruprocS=vyruskS/visoS*100
moteruprocS=moteruskS/visoS*100
```

Kaltininku procentas Klaipedos mieste.

```
vyruskKl<-sum(na.omit(as.numeric(Klaipedam[,19]=="Vyras")))
moteruskKl<-sum(na.omit(as.numeric(Klaipedam[,19]=="Moteris")))
visoKl<-(vyruskKl+moteruskKl)
vyruprocKl=vyruskKl/visoKl*100
moteruprocKl=moteruskKl/visoKl*100
```

Kaltininku procentas Panevezio mieste.

```
vyruskP<-sum(na.omit(as.numeric(Panevezysm[,19]=="Vyras")))
moteruskP<-sum(na.omit(as.numeric(Panevezysm[,19]=="Moteris")))
visoP<-(vyruskP+moteruskP)
vyruprocP=vyruskP/visoP*100
moteruprocP=moteruskP/visoP*100
```

Tikrinsime hipoteze ar Vilniaus mieste eismo ivykiu kaltininkai dazniau yra Vyrai nei moterys.

```
Vilniusm <- rawdata[rawdata[,1] == "Vilniaus m. sav.",]
sum(Vilniusm[, "Lytis"]=="Vyras")
```

```
## [1] NA
```



## koreliacijos

```
kor1 <- kor[-c(1,2,3),-c(1,2,3,5,8)]
kor1[,4] <- as.numeric(levels(kor1[,4]))[kor1[,4]]
cor(kor1)
```

```
##                               Ivykiu_sk._1k.gyv. Gyventoju_tankumas
## Ivykiu_sk._1k.gyv.                1.0000000      0.8555710
## Gyventoju_tankumas                0.8555710      1.0000000
## Auto_sk_1k_gyv                   -0.4425242     -0.4381115
## Islaidos_keliams_.mln._EUR.       -0.4681452     -0.6019193
##                               Auto_sk_1k_gyv Islaidos_keliams_.mln._EUR.
## Ivykiu_sk._1k.gyv.              -0.4425242      -0.4681452
## Gyventoju_tankumas              -0.4381115      -0.6019193
## Auto_sk_1k_gyv                  1.0000000       0.3935301
## Islaidos_keliams_.mln._EUR.      0.3935301       1.0000000
```

Palyginti zuvusiu/suzeistu menesinius vidurkius tarp laikotarpio ir vietu. Ar skiriasi ivykiu skaicius nakti ir diena?