

Darbą atliko – Airida Lipskaitė, 6grupė

Studento. kn. Nr. 1811406

Eil.Nr. 76

<b>apklausa</b>	Pajamų ir gyvenimo sąlygų tyrimas 2011 m. (asmenų duomenys)
<b>Salis</b>	Austrija

	x
<b>Eil.Nr</b>	76
<b>Stud.kn.Nr.</b>	1811406
<b>kintamasis1</b>	Prekių ir paslaugų importas, mln. eur., Imports of goods and services, metiniai
<b>kintamasis2</b>	Gimusiųjų skaičius vyrų, Live births - males, tūkst
<b>kintamasis3</b>	Užimti gyventojai (Active employment) - 25-54m., moterų, metiniai, tūkst.
<b>kintamasis4</b>	Vidutinės ekvivivalentinės neto pajamos, 16-64m. Mean and median income before social transfers (pensions included in social transfers)

(Šilintelė paimta iš Duomenų paskirstymas: pakeitimai)

Paaiškinimas:

- R kodas rašomas Courier New (11) šriftu.
- Įprastas tekstas rašomas Times New Roman (12) šriftu.
- R output rašomas Lucida Console (10) šriftu, ruda spalva (jog būtų greičiau pastebimas)

1. Parsisiųskite nurodytus duomenis (duomenų rinkmena toliau vadinama „apklausa“). Sudarykite duomenų rinkinį (parašyti R kodą), kuriame būtų nemažiau penkių stulpelių (kintamųjų). Bent vienas kintamasis turi būti kategorinis, bei bent vienas intervalinis.

```
#Install

install.packages("readxl")
#Load
library("readxl")
#xls file
apklausal <- read_excel("PGS 2011 asmenys_rev.xls")
#RB090 - lytis, PL060 - Valandų, praleistų pagrindiniame darbe,
skaičius per savaitę, PL040 - Statusas pagrindiniame darbe, PL080
- Mėnesių skaičius, kai buvote bedarbiu, PY010G - Samdomojo darbo
pajamos(per metus)
#RB090, PL040 - kategoriniai kintamieji; PY010G - intervalinis;
PL060, PL080 - kiekybiniai diskretūs
library(dplyr)
apklausa <- select(apklausal, RB090, PL060, PL040, PL080, PY010G)
#atspausdinu pirmas 6 eilutes iš df apklausa
head(apklausa)

# A tibble: 6 x 5
  RB090 PL060 PL040 PL080 PY010G
  <dbl> <dbl> <dbl> <dbl> <dbl>
1     1     60     3     0 73946.
2     2     20     3     0 4589.
3     2    NA     3     8     0
4     1     40     3     0 10306.
5     2     36     3     0 9839.
6     2    NA    NA     0     0

#atspausdinu paskutines 6 eilutes iš df apklausa
tail(apklausa)

# A tibble: 6 x 5
  RB090 PL060 PL040 PL080 PY010G
  <dbl> <dbl> <dbl> <dbl> <dbl>
1     1     40     3     0 39474.
2     2     20     3     0 74361.
3     1     40     3     0 9629.
4     2     40     3     0     0
5     2    NA    NA    NA    NA
6     1    NA    NA    NA    NA

apklausa$RB090 <- factor(apklausa$RB090)
apklausa$PL040 <- factor(apklausa$PL040)
apklausa$PL060 <- factor(apklausa$PL060)
apklausa$PL080 <- factor(apklausa$PL080)
```

---

2. Parsisiųskite ne mažiau kaip 15-os paskutiniųjų metų nurodytų rodiklių duomenis iš pasirinktų šaltinių (Eurostat, Pasaulio banko, Tarptautinio valiutos fondo, Lietuvos statistikos departamento ir t.t.). Sudarykite duomenų rinkinį (parašyti R kodą), kuriame būtų laiko stulpelis (METAI) ir rodiklių stulpeliai (duomenų rinkmena toliau vadinama „macro“).

```
#csv file
macro <- read.csv("macro.csv")
library(dplyr)
#atspausdinu pirmas 6 eilutes iš df macro
head(macro)
  Laikas kintamasis1 kintamasis2 kintamasis3 kintamasis4
1  2004    106451.8      40540      1401      15389
2  2005    115385.8      39878      1419      16523
3  2006    126538.1      39745      1440      16431
4  2007    136799.0      38940      1455      17301
5  2008    143428.7      40126      1469      19144
6  2009    120560.3      39072      1493      20148

#atspausdinu paskutines 6 eilutes iš df macro
tail(macro)
  Laikas kintamasis1 kintamasis2 kintamasis3 kintamasis4
10  2013    163979.9      40953      1544      21297
11  2014    166962.7      42162      1550      22792
12  2015    169856.4      43604      1552      22782
13  2016    173710.5      45051      1569      22610
14  2017    187871.4      45253      1571      24104
15  2018    200702.9         NA      1563      24448
```

---

### 3. Rinkmenoje „apklausa“

- a. Kategoriniams kintamiesiems sudarykite dažnių bei santykinų dažnių lenteles ir išbrėžkite stulpelių diagramą.

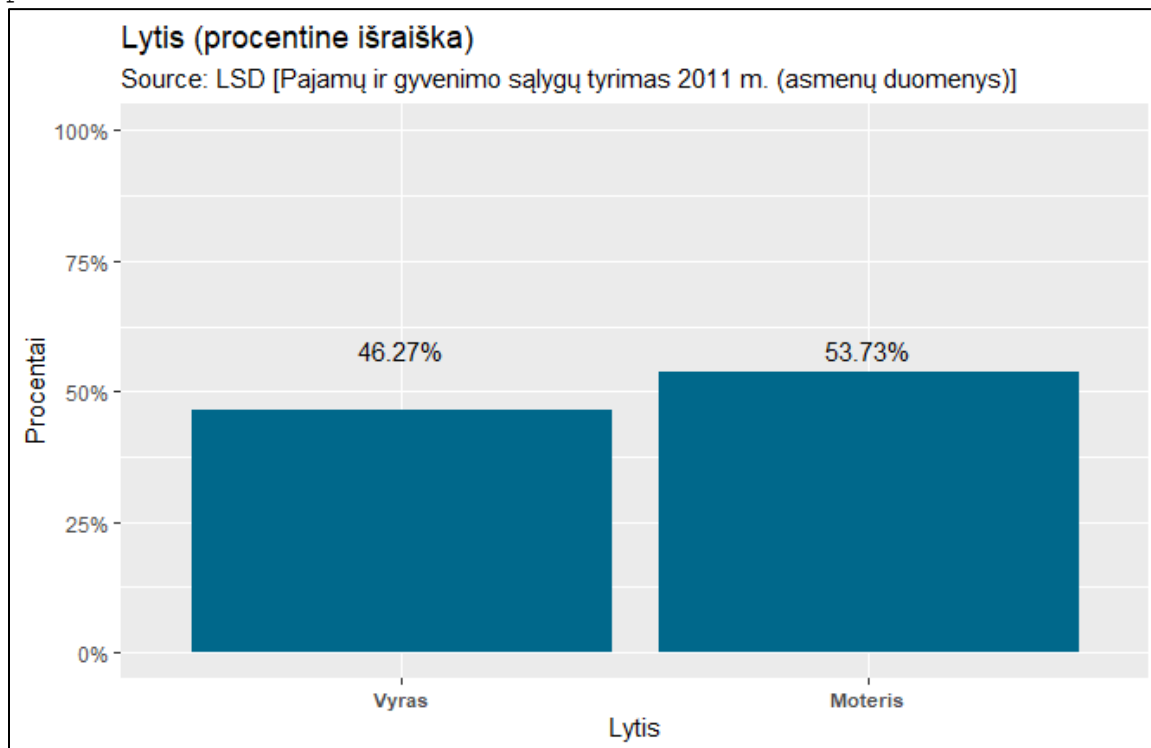
```
#####3_uzdotis#####
#####a_dalis#####
##
##dažnių lentelės##
#dažnių lentelė kintamajam RB090 - lytis.
table(apklausa$RB090)
  1      2
5780 6712
#dažnių lentelė kintamajam PL040 - Statusas pagrindiniame darbe
table(apklausa$PL040)
  1      2      3      4
128 395 9052 136
##santykinų dažnių lentelės##
#santykinų dažnių lentelė kintamajam RB090 - lytis
prop.table(table(apklausa$RB090))
  1      2
0.4626961 0.5373039

#santykinų dažnių lentelė kintamajam PL040 - Statusas pagrindiniame
darbe
```

```

prop.table(table(apklausa$PL040))
      1      2      3      4
0.01318093 0.04067552 0.93213881 0.01400474
##
#stulpelių diagrama##
library(ggplot2)
#stulpelių diagrama kinatamajam B090 - lytis
#paaiškinimas: Lytis 1 - Vyras, 2 - Moteris
p <- ggplot(apklausa, aes(RB090)) +
  geom_bar(aes(y = (..count..)/sum(..count..)), fill="deepskyblue4")+
  scale_y_continuous(labels=scales::percent)+
  scale_x_discrete(labels=c("Vyras", "Moteris"))+
  labs(title = "Lytis (procentine išraiška)",
        subtitle="Source: LSD [Pajamų ir gyvenimo sąlygų tyrimas
2011 m. (asmenų duomenys)]",
        x="Lytis",
        y="Procentai")+
  theme(axis.text.x = element_text(face="bold", size=8))+
  geom_text(aes(label=scales::percent(..count../sum(..count..)))
,
        stat='count',position=position_fill(vjust=0.58))
p

```



```

#stulpelių diagrama kinatamajam PL040 - Statusas pagrindiniame darbe
#Paaiškinimas: Statusas pagrindiniame darbe: 1 - Savarankiškai
dirbantis ir turintis samdomųjų darbuotojų asmuo, 2 - Savarankiškai
dirbantis be samdomųjų darbuotojų asmuo, 3 - Samdomas darbuotojas, 4 -
Dirbantis be atlyginimo šeimos versle, asmeniniame žemės ūkyje asmuo

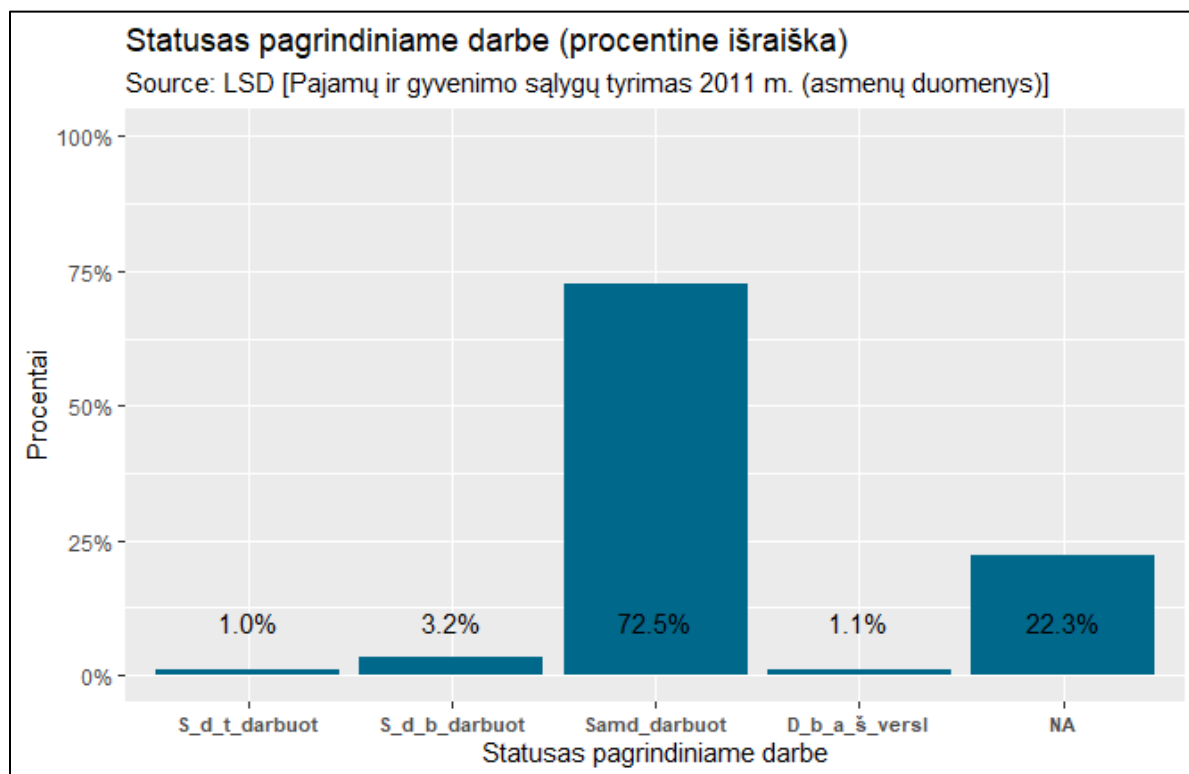
```

```

r <- ggplot(apklausa, aes(PL040)) +
  geom_bar(aes(y = (..count..)/sum(..count..)),
fill="deepskyblue4") +
  scale_y_continuous(labels=scales::percent)+
  scale_x_discrete(labels=c("S_d_t_darbuot", "S_d_b_darbuot",
"Samd_darbuot", "D_b_a_š_versl"))+
  labs(title = "Statusas pagrindiniame darbe (procentine
išraiška)",
  subtitle="Source: LSD [Pajamų ir gyvenimo sąlygų tyrimas
2011 m. (asmenų duomenys)]",
  x="Statusas pagrindiniame darbe",
  y="Procentai")+
  theme(axis.text.x = element_text(face="bold", size=8))+

geom_text(aes(label=scales::percent(..count../sum(..count..)),
  stat='count',position=position_fill(vjust=0.1))
r

```



b. Intervaliniams kintamiesiems apskaičiuokite padėties, sklaidos bei formos skaitines charakteristikas pagal kategorinio kintamojo(-ų) pjūvį(-ius).

```

#####b_dalis#####
dm<-select(apklausa, PY010G)
#Penkiaskaitė suvestinė/pagrindinės padėties charakteristikos
summary(dm)

```

```

      PY010G
Min.   :    0
1st Qu.:    0
Median :    0
Mean   : 8548
3rd Qu.:13370
Max.   :74872
NA's   :1464
#Papildomos sklaidos charakteristikos
summarise_all(dm,list(Stand_Nuok = sd, Dispersija = var, MAD =
mad),na.rm=T)
# A tibble: 1 x 3
  Stand_Nuok Dispersija   MAD
  <dbl>      <dbl> <dbl>
1  13854. 191941585.     0
#papildomos charakteristikos (asimetrijos koef. ir ekscesas)
install.packages("psych")
library(psych)
describe(data.frame(dm),na.rm=T)
  vars      n    mean      sd median trimmed mad min      max    range skew
x1     1 11028 8547.78 13854.3      0 5459.19   0   0 74871.88 74871.88 2.08
  kurtosis      se
x1      4.75 131.93
#Charakteristikos (pjūvis -- RB090 (Lytis))
#####
dma<-select(apklausa, RB090, PY010G)
dma %>% group_by(RB090)%>% summarise_all(list(Stand_Nuok = sd, Dispers
ija = var),na.rm=T)
# A tibble: 2 x 3
  RB090 Stand_Nuok Dispersija
  <dbl>      <dbl>      <dbl>
1     1      14481. 209706367.
2     2      13291. 176645591.
#Charakteristikos (pjūvis -- PL040 (Statusas pagrindiniame darbe))
#####
ma<-select(apklausa, PL040, PY010G)
ma %>% group_by(PL040)%>% summarise_all(list(Stand_Nuok = sd, Dispersi
ja = var),na.rm=T)
# A tibble: 5 x 3
  PL040 Stand_Nuok Dispersija
  <dbl>      <dbl>      <dbl>
1     1      10881. 118390317.
2     2       4420. 19536574.
3     3      14607. 213363560.
4     4       3853. 14843678.
5    NA       4097. 16787850.

```

c. Išbrėžkite pasirinktų rodiklių stačiakampę diagramą bei histogramą pagal kategorinio kintamojo (-ų) pjūvį(-ius).

```

#####C_dalis#####
#Pirma pasirenku du kintamuosius (šiuo atveju RB090 ir PY010G)
#Tuomet sukuriau nauja df pavadinimu - algos_pagal_lyti_vyru ir išfiltruoju PY010G, pagal RB090 1 pjūvį (Vyrai)
library(dplyr)
algos_pagal_lyti_vyru<-dma %>%
  filter (RB090=="1")
#Tuomet randu pagrindines padėties charakteristikas

```

```

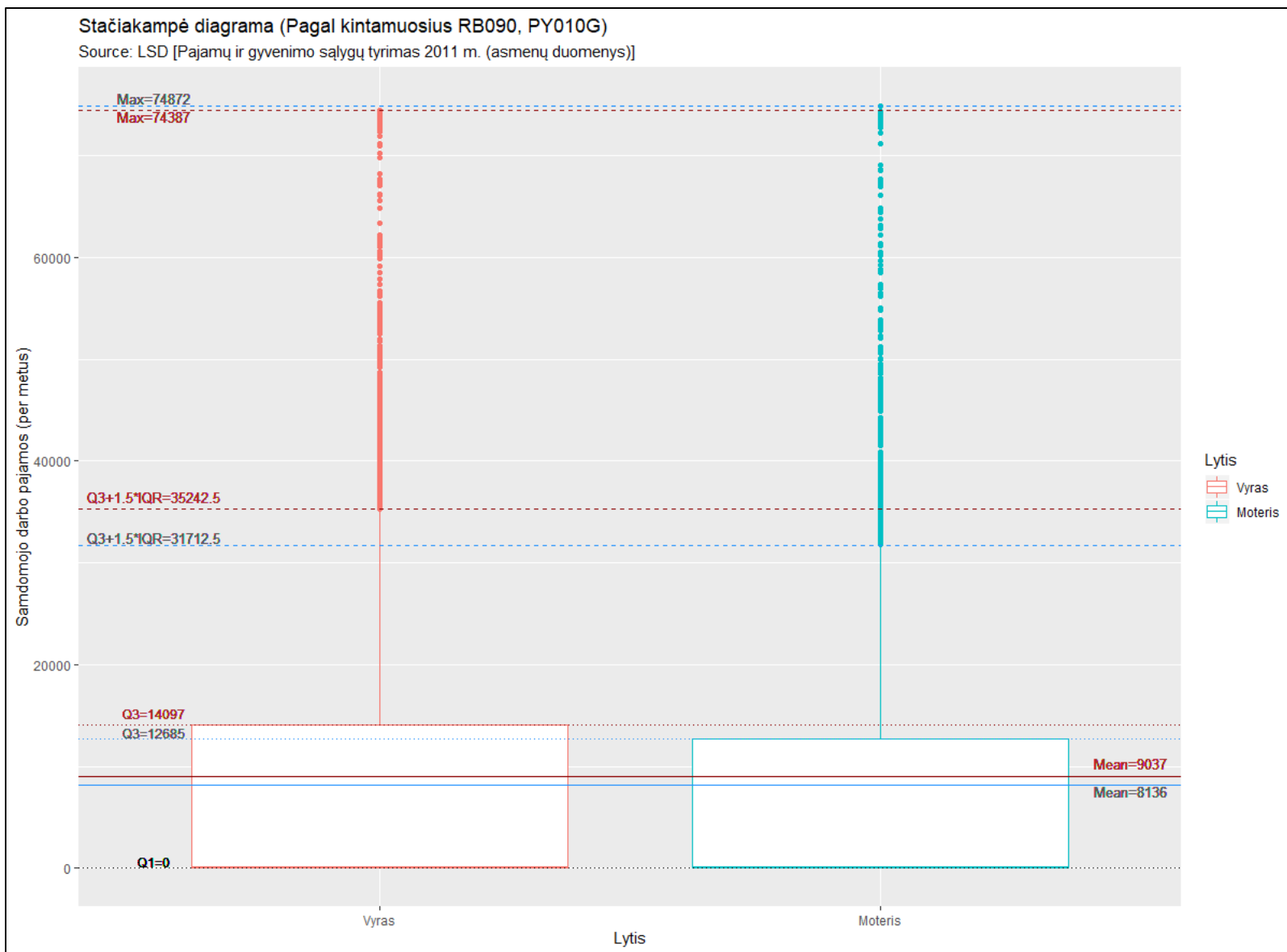
summary(algos_pagal_lyti_vyru, algos_pagal_lyti_vyru$PY010G)
  RB090      PY010G
Min.   :1    Min.   :    0
1st Qu.:1    1st Qu.:    0
Median :1    Median :    0
Mean   :1    Mean   : 9037
3rd Qu.:1    3rd Qu.:14097
Max.   :1    Max.   :74387
      NA's   :739
#Tada viską pakartoju pasirinkusi 2 RB090 pjuveni (moterys)
algos_pagal_lyti_mot<-dma %>%
  filter (RB090=="2")
summary(algos_pagal_lyti_mot, algos_pagal_lyti_mot$PY010G)
  RB090      PY010G
Min.   :2    Min.   :    0
1st Qu.:2    1st Qu.:    0
Median :2    Median :    0
Mean   :2    Mean   : 8136
3rd Qu.:2    3rd Qu.:12685
Max.   :2    Max.   :74872
      NA's   :725
# stačiakampė diagrama intervaliniam kintamajam - PY010G, pagal RB090
pjuvius
#Paaiškinimas PY010G - Samdomojo darbo pajamos(per metus), RB090 - lytis
is
ggplot(apklausa, aes(x=RB090, y=PY010G,color=as.factor(RB090)))+
  labs(x="Lytis", y="Samdomojo darbo pajamos (per metus)",color=
"Lytis", subtitle="Source: LSD [Pajamų ir gyvenimo sąlygų tyrimas 2011
m. (asmenų duomenys)]")+
  geom_boxplot()+
  ggtitle("Stačiakampė diagrama (Pagal kintamuosius RB090, PY010
G)")+
  scale_x_discrete(labels = c('Vyras','Moteris'))+
  geom_hline(yintercept=35242.5, linetype="dashed", color = "red
4")+
  geom_text(size=3.5, aes(0.55,35242.5,label = "Q3+1.5*IQR=35242.5", v
just = -0.6), colour="brown")+
  geom_hline(yintercept=31712.5, linetype="dashed", color = "dodgerblue
e")+
  geom_text(size=3.5, aes(0.55,31712.5,label = "Q3+1.5*IQR=31712.5", v
just = -0.15),colour="gray34")+
  geom_hline(yintercept=74387, linetype="dashed", color = "red4")+
  geom_text(size=3.5, aes(0.55,74387,label = "Max=74387", vjust = 1),
colour="brown")+
  geom_hline(yintercept=74872, linetype="dashed", color = "dodgerblue"
)+
  geom_text(size=3.5, aes(0.55,74872,label = "Max=74872", vjust = -0.1
5),colour="gray34")+
  geom_hline(yintercept=14097, linetype="dotted", color = "red4")+
  geom_text(size=3.1, aes(0.55,14097,label = "Q3=14097", vjust = -0.6)
, colour="brown")+
  geom_hline(yintercept=12685, linetype="dotted", color = "dodgerblue"
)+
  geom_text(size=3.1, aes(0.55,12685,label = "Q3=12685", vjust = -0.15
),colour="gray34")+

```

```

geom_hline(yintercept=0, linetype="dotted", color = "black")+
geom_text(size=3.1, aes(0.55,0,label = "Q1=0", vjust = -0.15),colour
="black")+
geom_hline(yintercept=9037, linetype="solid", color = "red4")+
geom_text(size=3.5, aes(2.5,9037,label = "Mean=9037", vjust = -0.6),
colour="brown")+
geom_hline(yintercept=8136, linetype="solid", color = "dodgerblue")+
geom_text(size=3.5, aes(2.5,8136,label = "Mean=8136", vjust = 1),col
our="gray34")+
scale_fill_discrete(name="Lytis",
breaks=c("1", "2"),
labels=c("Vyras", "Moteris"))

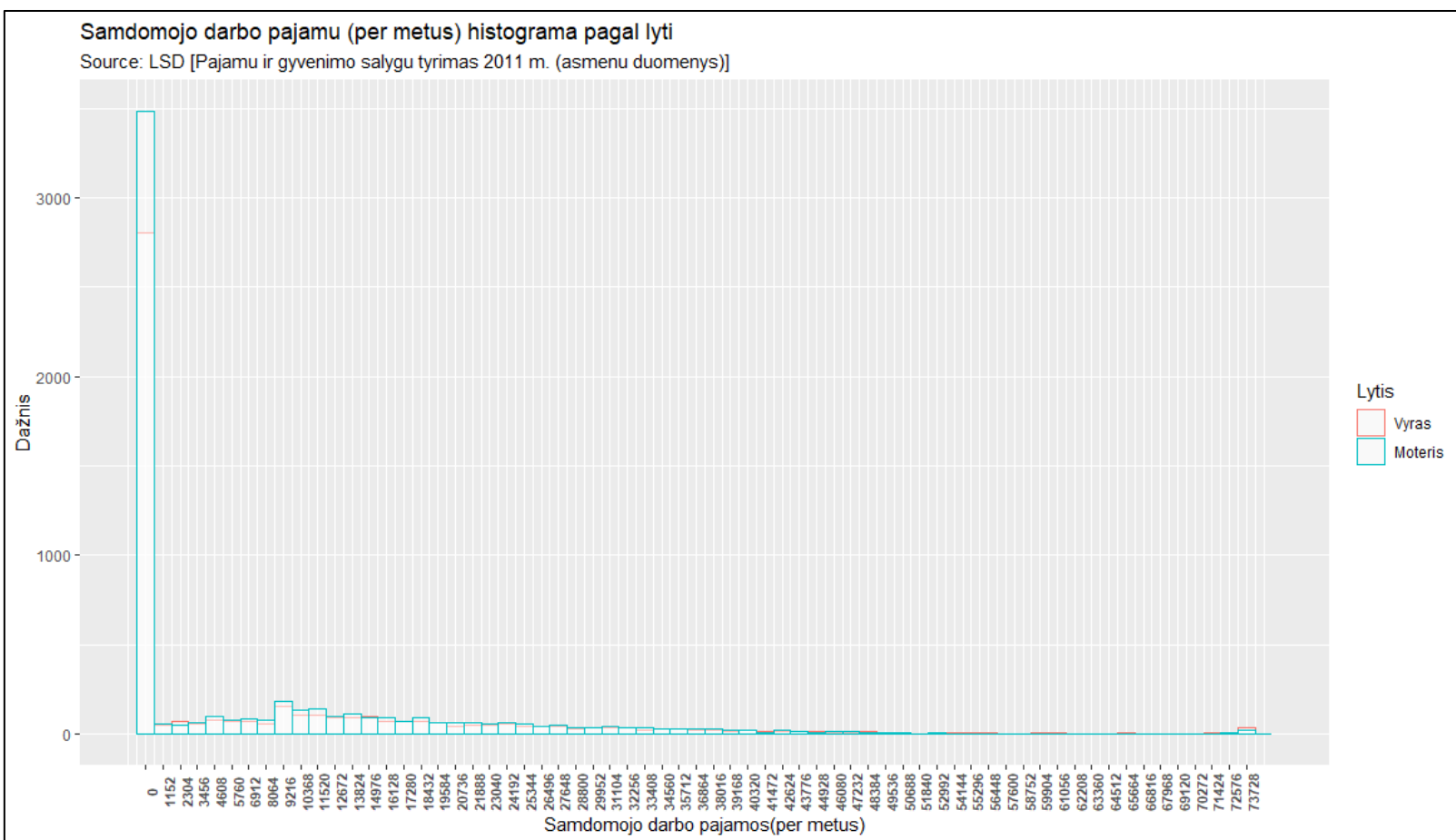
```





```
#Histograma PY010G(Samdomojo darbo pajamos(per metus)) pagal pjūvį
RB090(lytis)
apklausa$RB090 <- as.factor(apklausa$RB090)
levels(apklausa$RB090)<-c("Vyras","Moteris")

library(ggplot2)
ggplot(apklausa, aes(x=PY010G, group=RB090, color=RB090)) +
  geom_histogram(fill="white", binwidth=bw, alpha=0.5,
position="identity")+
  labs(x="Samdomojo darbo pajamos(per
metus)", y="Dažnis", group="Lytis", color="Lytis", subtitle="Source: LSD
[Pajamų ir gyvenimo sąlygų tyrimas 2011 m. (asmenų duomenys)]")+
  ggtitle("Samdomojo darbo pajamų (per metus) histograma pagal lytį")+
  scale_x_continuous(breaks=seq(0,74872, bw), labels = seq(0,74872,
bw)) +
  theme(axis.text.x = element_text(face="bold", size=8, angle=90))
```



```
(#Kad nustatytčiau bindwidth optimalų dydį naudojausi šią taisyklę
h=2×IQR×n-1/3)
```

```

#Pasirenku du kintamuosius (šiuo atveju PL040 ir PY010G)
#PL040 - Statusas pagrindiniame darbe
#Tuomet sukuriu nauja df pavadinimu - stat_sav_su_darbuot ir išfiltruojau
PY010G, pagal PL040 1 pjūvį (Savarankiškai dirbantis ir turintis samdomųjų
darbuotojų asmuo)
library(dplyr)
stat_sav_su_darbuot<-ma %>%
  filter (PL040=="1")
#Tuomet randu pagrindines padėties charakteristikas
summary(stat_sav_su_darbuot, stat_sav_su_darbuot$PY010G)
      PL040      PY010G
Min.   :1    Min.   :    0
1st Qu.:1    1st Qu.:    0
Median :1    Median :    0
Mean   :1    Mean   : 4417
3rd Qu.:1    3rd Qu.:    0
Max.   :1    Max.   :73968
#Tada viską pakartoju pasirinkusi 2 PL040 pjūvį (Savarankiškai dirbant
is be samdomųjų darbuotojų asmuo)
stat_sav_be_darbuot<-ma %>%
  filter (PL040=="2")
summary(stat_sav_be_darbuot, stat_sav_be_darbuot$PY010G)
      PL040      PY010G
Min.   :2    Min.   :    0
1st Qu.:2    1st Qu.:    0
Median :2    Median :    0
Mean   :2    Mean   : 1204
3rd Qu.:2    3rd Qu.:    0
Max.   :2    Max.   :31323
#Pasirenku 3 PL040 pjūvį (Samdomas darbuotojas) ir pakartoju prieš tai
atliktus veiksmus
samdomas_darbuotojas<-ma %>%
  filter (PL040=="3")
summary(samdomas_darbuotojas, samdomas_darbuotojas$PY010G)
      PL040      PY010G
Min.   :3    Min.   :    0
1st Qu.:3    1st Qu.:    0
Median :3    Median : 1597
Mean   :3    Mean   :10194
3rd Qu.:3    3rd Qu.:16002
Max.   :3    Max.   :74872
#Pasirenku 4 PL040 pjūvį (Dirbantis be atlyginimo šeimos versle, asmen
iniame žemės ūkyje asmuo) ir pakartoju prieš tai atliktus veiksmus
dirb_be_atlyg<-ma %>%
  filter (PL040=="4")
summary(dirb_be_atlyg, dirb_be_atlyg$PY010G)
library(ggplot2)

## stačiakampė diagrama intervaliniam kintamajam - PY010G, pagal PL040
pjūvius
ggplot(na.omit(apklausa), aes(x=PL040, y=PY010G,color=as.factor(PL040)
))+
  labs(x="Statusas pagrindiniame darbe", y="Samdomojo darbo pajamos (p
er metus)",color="Statusas pagrindiniame darbe", subtitle="Source: LSD
[Pajamų ir gyvenimo sąlygų tyrimas 2011 m. (asmenų duomenys)]")+
  geom_boxplot()+
  ggtitle("Stačiakampė diagrama (Pagal kintamuosius PL040, PY010G)")

```

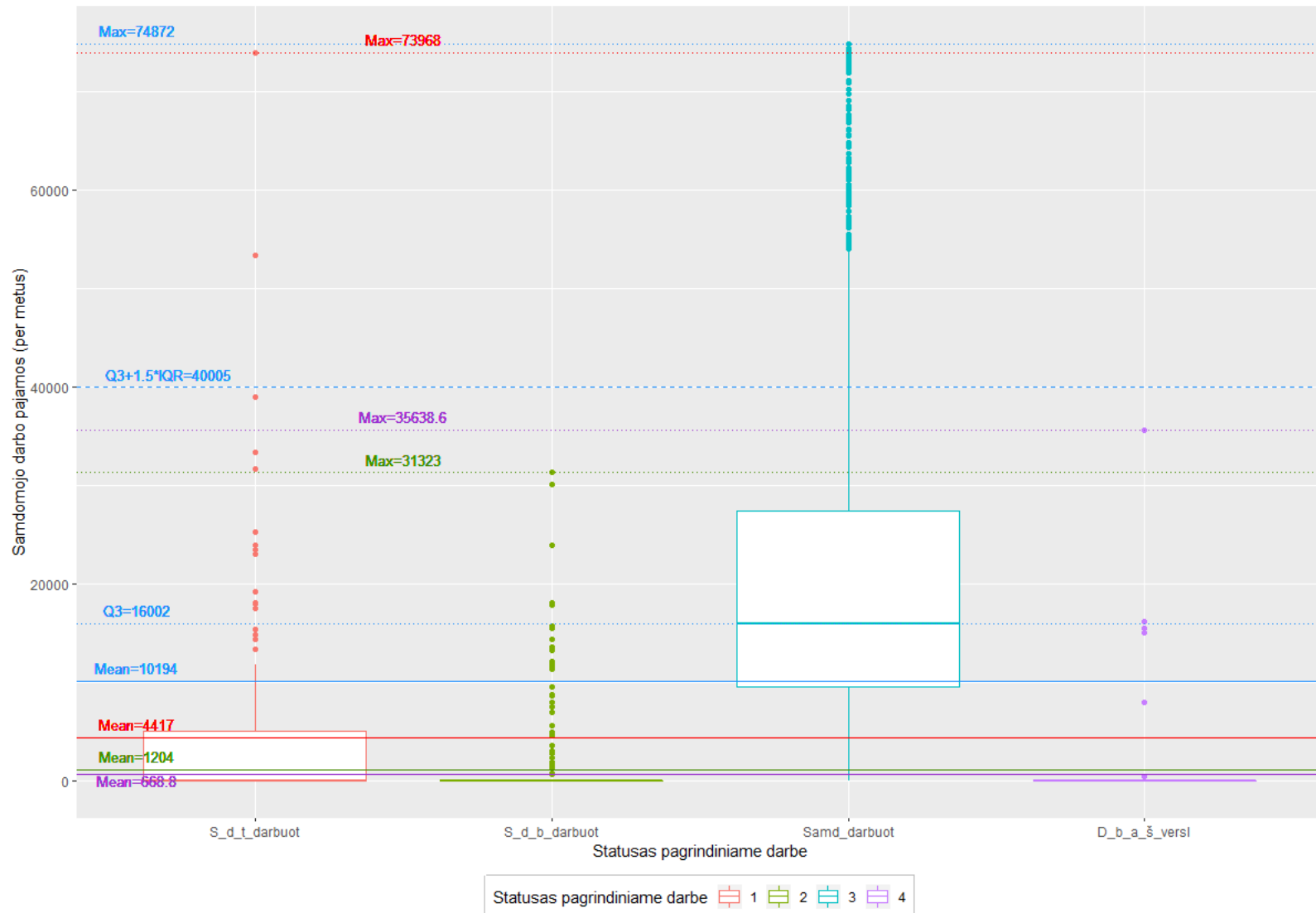
```

scale_x_discrete(labels = c('S_d_t_darbuot','S_d_b_darbuot', "Samd_d
arbuot","D_b_a_š_versl"))+
theme(legend.position="bottom",
      legend.box.background = element_rect(color="gray", size=1))+
scale_fill_discrete(name="Statusas pagrindiniame darbe",
                    breaks=c("1", "2", "3", "4"),
                    labels=c("S_d_t_darbuot", "S_d_b_darbuot", "Samd
_darbuot", "D_b_a_š_versl"))+
geom_hline(yintercept=10194, linetype="solid", color = "dodgerblue")
+
geom_text(size=3.5, aes(0.6,10194,label = "Mean=10194", vjust = -0.6
), colour="dodgerblue")+
geom_hline(yintercept=16002, linetype="dotted", color = "dodgerblue"
)+
geom_text(size=3.5, aes(0.6,16002,label = "Q3=16002", vjust = -0.6),
colour="dodgerblue")+
geom_hline(yintercept=74872, linetype="dotted", color = "dodgerblue"
)+
geom_text(size=3.5, aes(0.6,74872,label = "Max=74872", vjust = -0.6)
, colour="dodgerblue")+
geom_hline(yintercept=40005, linetype="dashed", color = "dodgerblue"
)+
geom_text(size=3.5, aes(0.71,40005,label = "Q3+1.5*IQR=40005", vjust
= -0.6), colour="dodgerblue")+
geom_hline(yintercept=4417, linetype="solid", color = "red")+
geom_text(size=3.5, aes(0.6,4417,label = "Mean=4417", vjust = -0.6),
colour="red")+
geom_hline(yintercept=73968, linetype="dotted", color = "red")+
geom_text(size=3.5, aes(1.5,73968,label = "Max=73968", vjust = -0.6)
, colour="red")+
geom_hline(yintercept=1204, linetype="solid", color = "chartreuse4")
+
geom_text(size=3.5, aes(0.6,1204,label = "Mean=1204", vjust = -0.6),
colour="chartreuse4")+
geom_hline(yintercept=31323, linetype="dotted", color = "chartreuse4
")+
geom_text(size=3.5, aes(1.5,31323,label = "Max=31323", vjust = -0.6)
, colour="chartreuse4")+
geom_hline(yintercept=668.8, linetype="solid", color = "darkorchid3"
)+
geom_text(size=3.5, aes(0.6,668.8,label = "Mean=668.8", vjust = 1),
colour="darkorchid3")+
geom_hline(yintercept=35638.6, linetype="dotted", color = "darkorchi
d3")+
geom_text(size=3.5, aes(1.5,35638.6,label = "Max=35638.6", vjust = -
0.6), colour="darkorchid3")

```

# Stačiakampė diagrama (Pagal kintamuosius PL040, PY010G)

Source: LSD [Pajamų ir gyvenimo sąlygų tyrimas 2011 m. (asmenų duomenys)]



#Šiuo atveju stačiakampėje histogramoje nėra atvaizduotų Min, Q1 reikšmių, nes panaudojus summary f-ja skirtingiems PL040 pjūviams paaiškėjo, kad tos reikšmės lygios 0

#taip pat pjūviams S\_d\_t\_darbuot, S\_d\_b\_darbuot ir D\_b\_a\_š\_versl panaudojus summary f-ja, tapo aišku, jog Median ir Q3 lygūs 0

#Šiuo atveju Median turėjo tik trečiasis pjūvis (Samd\_darbuot) ir Median=1597 (visos kitos pagr. summary reikšmės yra atvaizduotos grafike)

\*Legendoje 1 atitinka - S\_d\_t\_darbuot, 2 atitinka - S\_d\_b\_darbuot, 3 atitinka - Samd\_darbuot, 4 atitinka - D\_b\_a\_š\_versl

#Histograma PY010G(Samdomojo darbo pajamos(per metus)) pagal pjūvį PL040(statusas pagrindiniame darbe)

```
apklausa$PL040 <- as.factor(apklausa$PL040)
```

```
levels(apklausa$PL040)<-c('S_d_t_darbuot','S_d_b_darbuot',
"Samd_darbuot","D_b_a_š_versl")
```

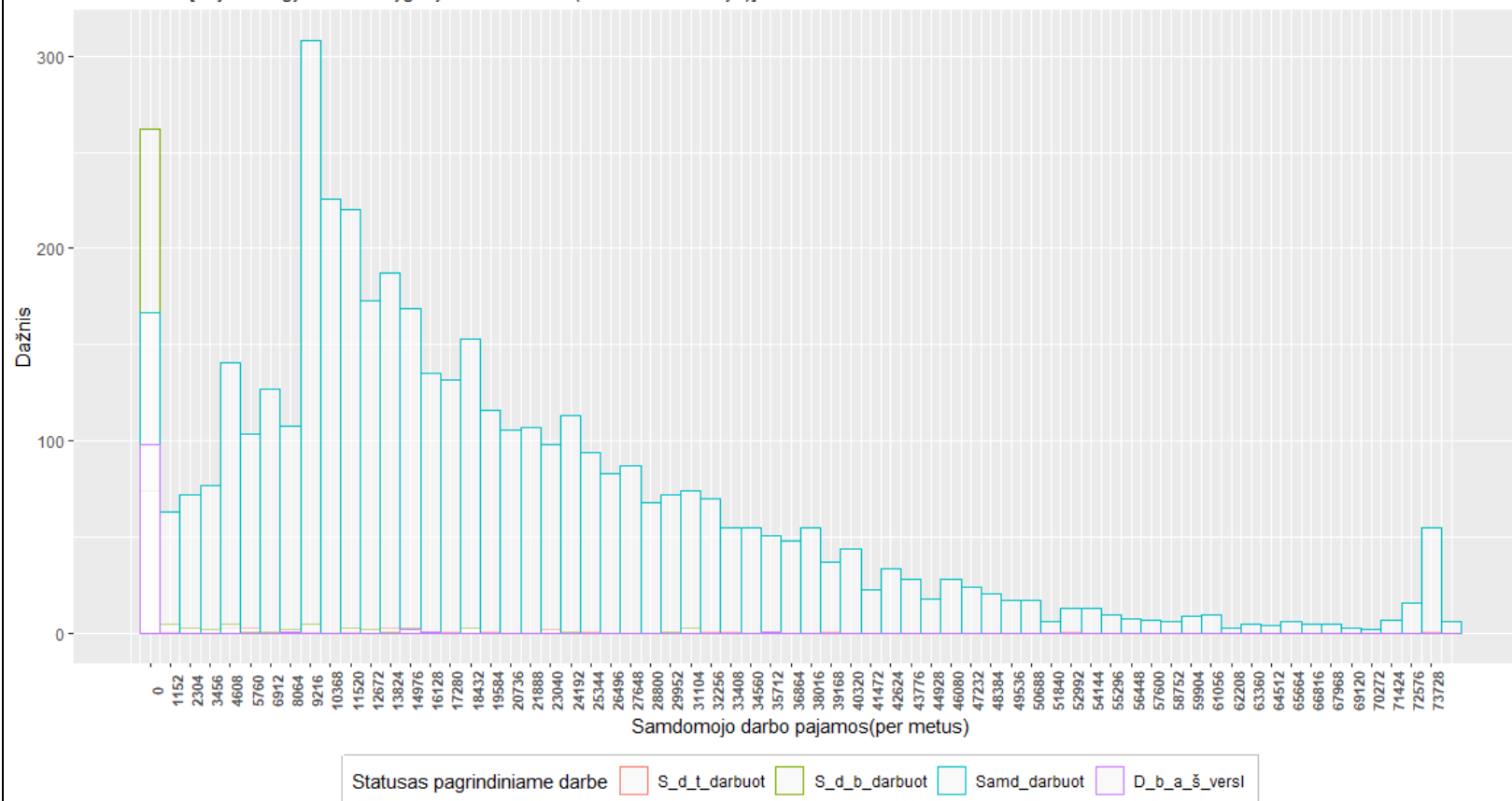
```
#Kad nustatyčiau bindwidth optimalų dydį naudosiu šią taisyklę
h=2×IQR×n-1/3

bw <- round(2 * IQR(apklausa$PY010G, na.rm=T) /
length(apklausa$PY010G)^(1/3))

ggplot(na.omit(apklausa), aes(x=PY010G, group=PL040, color=PL040)) +
geom_histogram(fill="white",binwidth=bw, alpha=0.5,
position="identity")+
  labs(x="Samdomojo darbo pajamos(per
metus)",y="Dažnis",group="Statusas pagrindiniame
darbe",color="Statusas pagrindiniame darbe", subtitle="Source: LSD
[Pajamų ir gyvenimo sąlygų tyrimas 2011 m. (asmenų duomenys)]")+
  ggtitle("Samdomojo darbo pajamų (per metus) histograma pagal PL040
pjuvius")+
  scale_x_discrete(labels = c('S_d_t_darbuot','S_d_b_darbuot',
"Samd_darbuot","D_b_a_š_versl"))+
  theme(legend.position="bottom",
        legend.box.background = element_rect(color="gray", size=1))+
  scale_x_continuous(breaks=seq(0,74872, bw),labels = seq(0,74872,
bw))+theme(axis.text.x = element_text(face="bold", size=8, angle=90))
```

### Samdomojo darbo pajamų (per metus) histograma pagal PL040 pjuvius

Source: LSD [Pajamų ir gyvenimo sąlygų tyrimas 2011 m. (asmenų duomenys)]



4. Apibūdinkite „macro“ kintamuosius; apskaičiuokite padėties, sklaidos bei formos skaitines charakteristikas; pateikite gautų charakteristikų interpretaciją bei ekonominį pagrindimą; Išbrėžkite pasirinktų rodiklių linijos grafiką bei histogramą.

**„macro“ kintamųjų apibūdinimas:**

- **Kintamasis1** - Prekių ir paslaugų importas, mln., eur., Imports of goods and services, metiniai.

(Šaltinis: Eurostat [nama\_10\_gdp].

Duomenų lentelės pavadinimas: *GDP and main components (output, expenditure and income)*.

Pasirinktas laikotarpis (TIME) - 2004-2018 m.

Šalis (GEO) – Austrija (AT).

Nacionalinių sąskaitų rodiklis (NA\_ITEM) – prekių ir paslaugų importas (P7)

Matavimo vienetai/valiuta (UNIT) – mln., eur. (CP\_MEUR)).

Apibrėžimas **Importas** - *Prekių, paslaugų, darbo ir kapitalo įvežimas (pirkimas) iš užsienio šalies* (Šaltinis: <http://zodynas.vz.lt/Importas>). Kintamojo1 atveju **importas** yra apibrėžtas kaip prekių ir paslaugų įvežimas iš užsienio šalies)

- **Kintamasis2** - Gimusiųjų vyrų skaičius, Live births - males, tūkst.

(Šaltinis: Eurostat [demo\_gind].

Duomenų lentelės pavadinimas: *Population change - Demographic balance and crude rates at national level*.

Pasirinktas laikotarpis (TIME) - 2004-2018 m.

Šalis (GEO) – Austrija (AT).

Demografinis indikatorius (INDIC\_DE) – Gimusiųjų vyrų skaičius (MLBIRTH)).

Apibrėžimas: **Gimusiųjų skaičius** – *skaičius kūdikių, kurie gimė gyvi* (Šaltinis: <https://www.collinsdictionary.com/dictionary/english/live-birth>)

- **Kintamasis3** - Užimti gyventojai (Active employment) - 25-54m., moterų, metiniai, tūkst.

(Šaltinis: Eurostat [lfsi\_emp\_a].

Duomenų lentelės pavadinimas: *Employment and activity by sex and age - annual data*.

Pasirinktas laikotarpis (TIME) - 2004-2018 m.

Šalis (GEO) – Austrija (AT).

Amžius (AGE) – 25-54 m. (Y25-54).

Užimtumo indikatorius (INDIC\_EM) – aktyvi populiacijos dalis (ACT).

Lytis (SEX) – moteris (F).

Matavimo vienetai/valiuta (UNIT) – tūkst. Moterų (THS\_PER)).

Apibrėžimas: **Užimti gyventojai** – *asmenys, dirbantys bet kokią darbą, gaunantys už jį darbo užmokestį pinigais ar natūra arba turintys pajamų ar pelno* (Šaltinis: <http://zodynas.vz.lt/uzimti-gyventojai>)

- **Kintamasis4** - Vidutinės ekvivalentinės neto pajamos, 16-64m. Mean and median income before social transfers (pensions included in social transfers).  
(Šaltinis: Eurostat [ilc\_di13].  
Duomenų lentelės pavadinimas: *Mean and median income before social transfers (pensions included in social transfers) by age and sex.*  
Pasirinktas laikotarpis (TIME) 2004-2018 m.  
Šalis (GEO) – Austrija (AT).  
Amžius (AGE) – 16-64 m. (Y16-64).  
Pajamų ir gyvenimo būklės indikatorius (INDIC\_IL) – Vidutinės grynosios pajamos (MEI\_E).  
Lytis (SEX) – vyrai (M).  
Matavimo vienetai/valiuta (UNIT) – eurai (EUR)).

Apibrėžimas: Neto pajamos (arba neto darbo užmokestis) - atlyginimas pinigais, apimantis tiesiogiai darbdavio darbuotojui mokamą pagrindinį darbo užmokestį ir papildomą uždarbį, atskaičius darbdavio mokamas socialinio draudimo įmokas ir gyventojų pajamų mokestį (Šaltinis: <http://zodynas.vz.lt/neto-darbo-uzmokestis>).

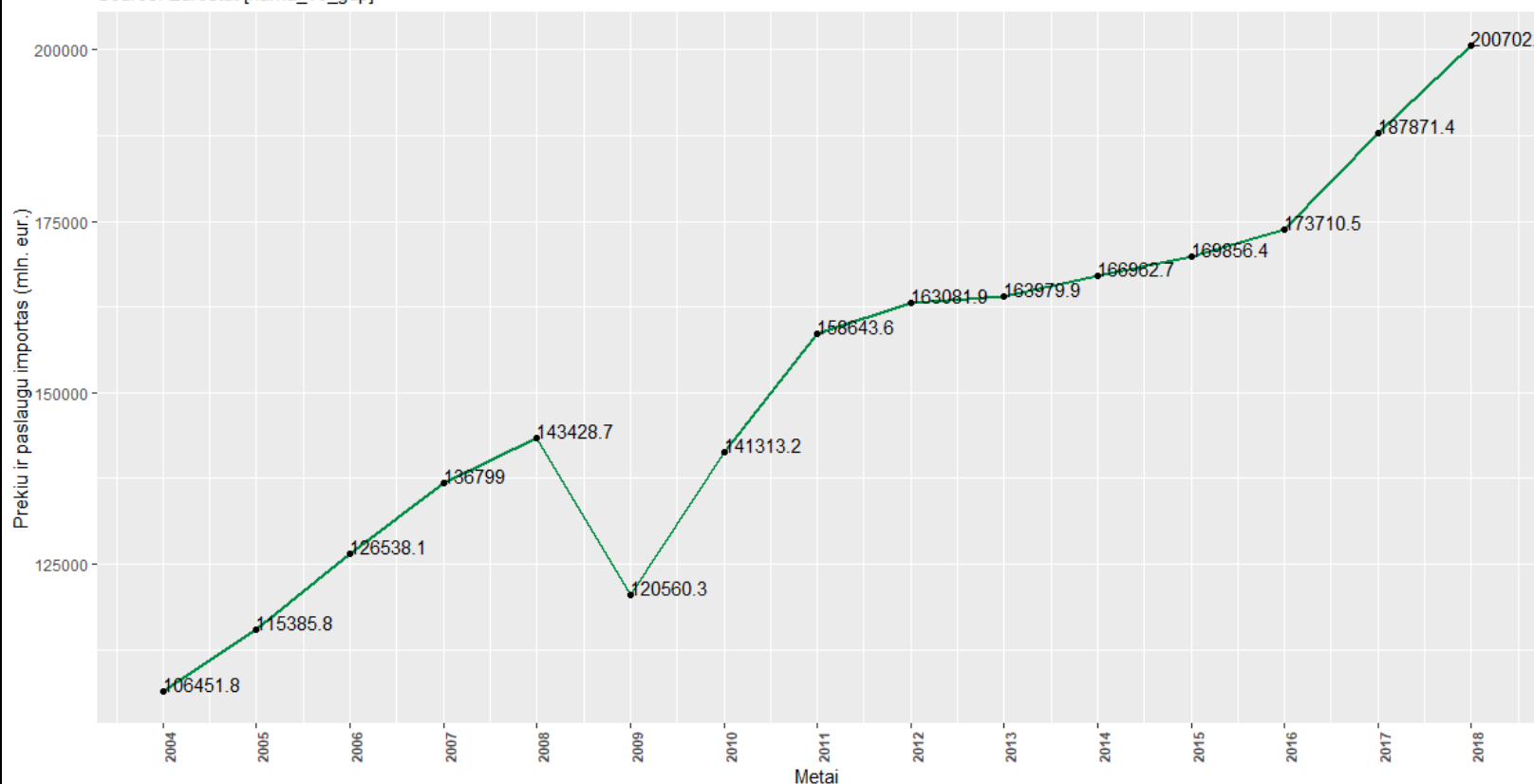
```
#Pirma pašalinu laiko stulpelį, nes tai ne intervalinis kintamasis
macro1<-select(macro, -Laikas)
#Pagrindinės padėties skaitinės charakteristikos
summary(macro1)
  kintamasis1      kintamasis2      kintamasis3      kintamasis4
Min.   :106452    Min.   :38940    Min.   :1401    Min.   :15389
1st Qu.:131669    1st Qu.:39940    1st Qu.:1462    1st Qu.:18223
Median :158644    Median :40468    Median :1516    Median :20511
Mean   :151686    Mean   :41205    Mean   :1505    Mean   :20350
3rd Qu.:168410    3rd Qu.:41860    3rd Qu.:1551    3rd Qu.:22696
Max.   :200703    Max.   :45253    Max.   :1571    Max.   :24448
      NA's      :1
#Papildomos sklaidos charakteristikos
summarise_all(macro1,list(Stand_Nuok = sd, Dispersija = var))
  kintamasis1_Stand_Nuok kintamasis2_Stand_Nuok kintamasis3_Stand_Nuok
1          27324.16              NaN              56.54387
  kintamasis4_Stand_Nuok kintamasis1_Dispersija kintamasis2_Dispersija
1          2869.126              746609869              NA
  kintamasis3_Dispersija kintamasis4_Dispersija
1          3197.21              8231884
#papildomos charakteristikos (asimetrijos koef. ir ekscesas)
library(psych)
describe(data.frame(macro1),na.rm=T)
  vars  n  mean  sd  median  trimmed  mad  min
kintamasis1 1 15 151685.75 27324.16 158643.6 151394.73 25694.05 106451.8
kintamasis2 2 14 41205.14 2050.48 40467.5 41056.58 972.59 38940.0
kintamasis3 3 15 1504.93 56.54 1516.0 1507.85 69.68 1401.0
kintamasis4 4 15 20349.60 2869.13 20511.0 20415.92 3366.98 15389.0
      max  range skew kurtosis  se
kintamasis1 200702.9 94251.1 -0.01 -1.18 7055.07
kintamasis2 45253.0 6313.0 0.90 -0.66 548.01
kintamasis3 1571.0 170.0 -0.44 -1.33 14.60
kintamasis4 24448.0 9059.0 -0.29 -1.30 740.81
#####
```

```
#####linijų_grafikas#####
library(ggplot2)

ggplot(data=macro, aes(x=Laikas, y=kintamasis1, label=kintamasis1))+
  geom_line(color="springgreen4", size=1)+
  geom_point()+
  geom_text(aes(label=kintamasis1), hjust=0, vjust=0)+
  scale_x_continuous(breaks=macro$Laikas, labels = macro$Laikas)+
  theme(axis.text.x = element_text(face="bold", size=8, angle=9
0)))+
  labs(x="Metai", y="Prekių ir paslaugų importas (mln. eur.)", s
ubtitle ="Source: Eurostat [nama_10_gdp]")+
  ggtitle("Prekių ir paslaugų importas (mln. eur.) 2004-2018 m.
")
```

**Prekių ir paslaugų importas (mln. eur.) 2004-2018 m.**

Source: Eurostat [nama\_10\_gdp]



### **Ekonominis pagrindimas remiantis kintamuoju1 (Prekių ir paslaugų importas, mln., eur)**

Taigi tiek suradę pagrindines skaitines padėties charakteristikas, tiek išbrėžią linijinį grafiką, matome, jog didžiausia kintamojo1 reikšmė buvo 200703 mln.e. (2018 m.), taip yra todėl, jog 2018 metais tiek pasaulinė, tiek ES (ir tuo pačiu Austrijos) ekonomika buvo susidūrusi su labai sparčiu ekonomikos augimu. Įprasta, jog augant ekonomikai, auga ir, prekių ir paslaugų importas iš užsienio šalių.

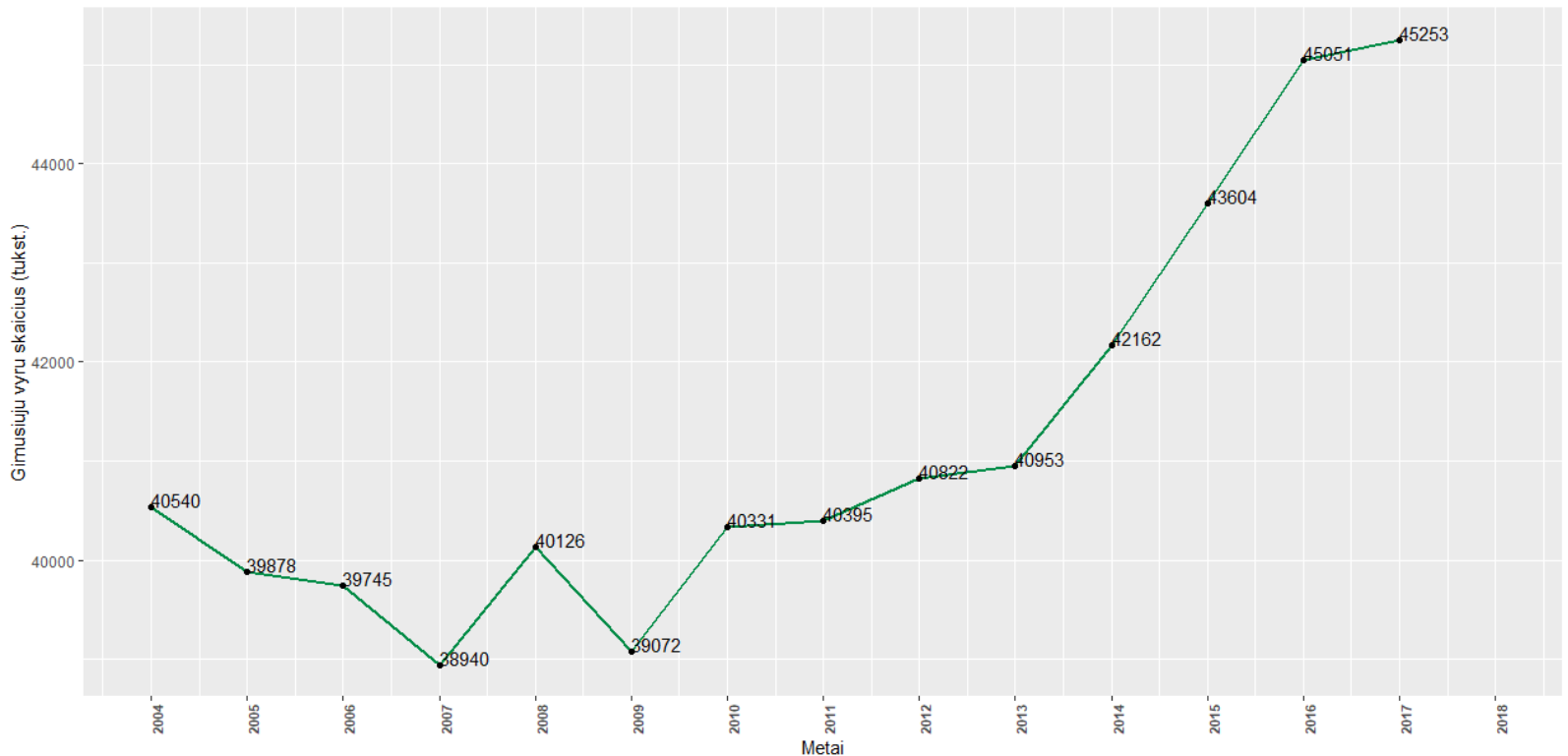


Minimali kintamojo1 reikšmė buvo 106452 mln. e. (2004 m.), sunku įvardinti pagrindinius veiksnius, kurie tai nulėmė, nes 2004 m. Austrijoje ekonominė situacija buvo labai gera (2004 m. Austrija buvo ketvirta pagal turtingumą šalis ES, ją pralenkė tik Liuksemburgas, Airija ir Nyderlandai.). Mano manymu, importas 2004 m. buvo pakankamai žemas, nes tuo metu ES ekonomika nebuvo pasiekusi tokio atvirumo lygio kaip dabar (t.y. šalių ekonomikos buvo kur kas uždaresnės, nelinkusios importuoti tiek daug prekių ir paslaugų). Taip pat iš linijų grafiko matome, jog 2008-2009 metų krizė turėjo nemenką įtaką prekių ir paslaugų importui, t.y. nuo 2009 m. importas sumažėjo beveik 1,2 karto palyginus su 2008 m. Taip yra todėl, jog recesijos metu, paprastai mažėja vartojimas, vartotojai tampa taupesni, nes bijo prarasti turimus darbus, nežino ko tikėtis ateityje. Todėl pasirinkdami taupymo metodą, jie nebegali sau leisti pirkti tiek daug prekių ir paslaugų, kaip anksčiau ir tokia vartotojų reakcija atsispindi ne tik bendrame prekių ir paslaugų vartojimo sumažėjime, bet ir prekių ir paslaugų importo duomenyse.

```
ggplot(data=macro, aes(x=Laikas, y=kintamasis2, label=kintamasis2))+
  geom_line(color="springgreen4", size=1)+
  geom_point()+
  geom_text(aes(label=kintamasis2), hjust=0, vjust=0)+
  scale_x_continuous(breaks=macro$Laikas, labels = macro$Laikas)+
  theme(axis.text.x = element_text(face="bold", size=8,
angle=90))+
  labs(x="Metai", y="Gimusiųjų vyrų skaičius (tūkst.)", subtitle
="Source: Eurostat [demo_gind]")+
  ggtitle("Gimusiųjų vyrų skaičius (tūkst.) 2004-2018 m.")
```

Gimusiųjų vyrų skaičius (tūkst.) 2004-2018 m.

Source: Eurostat [demo\_gind]



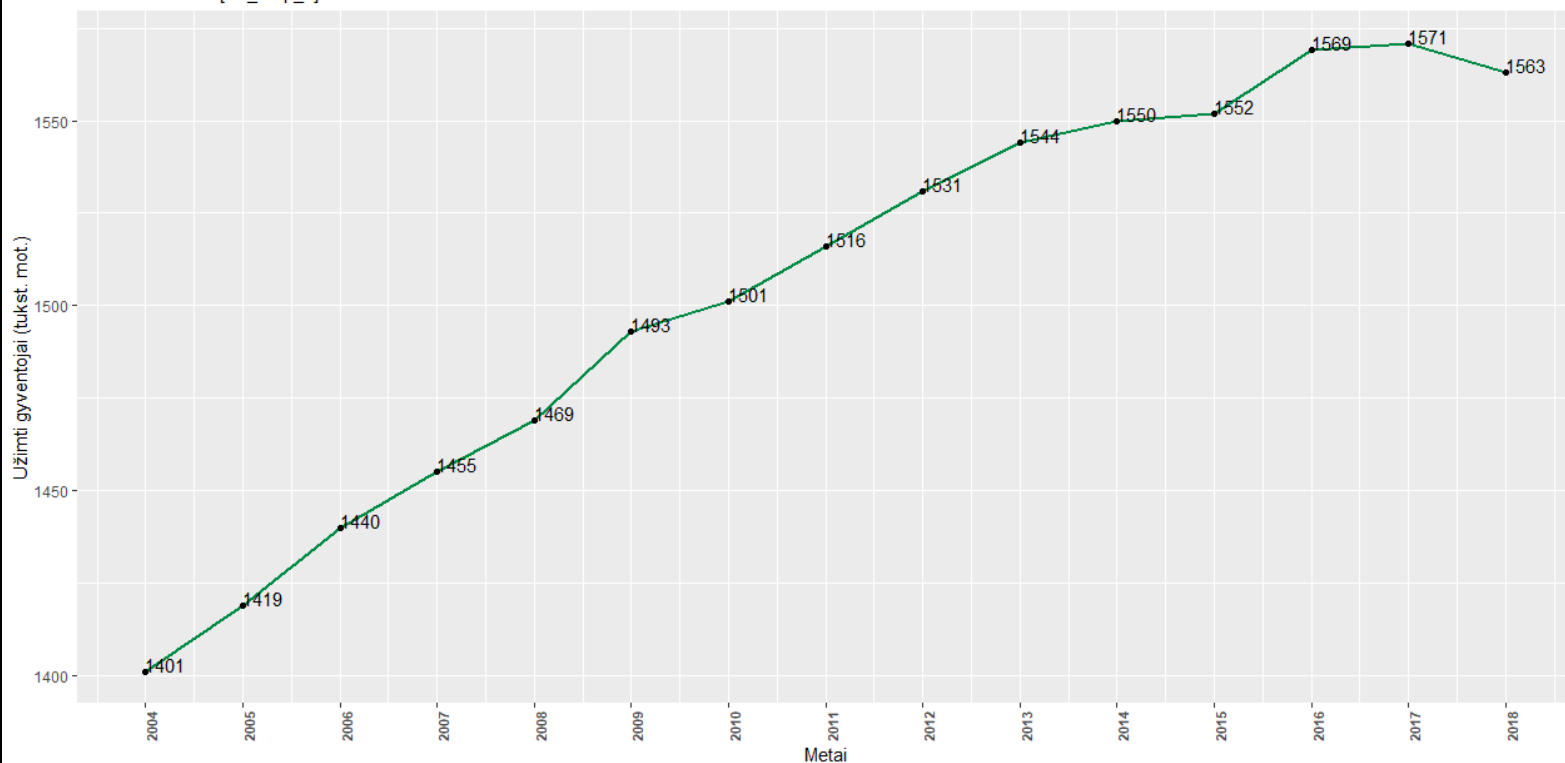
## **Ekonominis pagrindimas remiantis kintamuoju2 (Gimusiųjų vyrų skaičius (tūkst.))**

Taigi tiek suradę pagrindines skaitines padėties charakteristikas, tiek išbrėžią linijinį grafiką, matome, jog didžiausia kintamojo2 (Gimusiųjų vyrų skaičius (tūkst.)) reikšmė buvo 45253 tūkst. (2017 m.). O štai minimali kintamojo2 (Gimusiųjų vyrų skaičius (tūkst.)) reikšmė buvo 38940 (2007 m.). Sunku įvardinti, kas galėjo nulemti tokį gimusiųjų vyrų skaičių Austrijoje, tačiau galima daryti prielaidą, jog bėgant laikui, ekonomika auga, augant ekonomikai vis daugiau lėšų gali būti skiriama sveikatos apsaugai, ligoninių renovacijai, naujų technologijų diegimui (kuris gali sumažinti ir kūdikių mirtingumą), specialistų kvalifikacijos kėlimui. Tai paaiškinant faktą, kodėl nuo 2010 m. iki 2017 m. gimusiųjų vyrų skaičius Austrijoje tendencingai išaugo. Taip pat dažnu atveju padidėjęs gimusiųjų vaikų skaičius (nebūtinai tik vyrų) gali pasireikšti ir dėl valdžios padidintų transferų (šiuo atveju, galbūt Austrijos valdžia nusprendė padidinti paramą šeimoms kurios susilaukia daugiau vaikų, šeimoms, kurių prastesnė finansinė padėtis ir pan.). Be visų anksčiau paminėtų priežasčių populiacijos augimui įtakos turi ir individų požiūris, gal būti, kad individai susilaukia vaikų tuomet kai jaučia didesnę valstybės paramą, stiprų ekonomikos augimą (kurį ji tiesiogiai pajaučia augant jų darbo užmokesčiui). Taip pat individų polinkį susilaukti vaikų gali skatinti ir asmeninės priežastys, kurias lemia charakteris, temperamentas ir pan. Žinoma, be viso to individo polinkį susilaukti vaikų gali nulemti ir žiniasklaidos formuojama nuomonė

```
ggplot(data=macro, aes(x=Laikas, y=kintamasis3, label=kintamasis3))+
  geom_line(color="springgreen4", size=1)+
  geom_point()+
  geom_text(aes(label=kintamasis3), hjust=0, vjust=0)+
  scale_x_continuous(breaks=macro$Laikas, labels = macro$Laikas)+
  theme(axis.text.x = element_text(face="bold", size=8,
angle=90))+
  labs(x="Metai", y="Užimti gyventojai (tūkst. mot.)", subtitle
="Source: Eurostat [lfsi_emp_a]")+
  ggtitle("Užimti gyventojai (tūkst. mot.) 2004-2018 m.")
```

Užimti gyventojai (tūkst. mot.) 2004-2018 m.

Source: Eurostat [lfsi\_emp\_a]



### Ekonominis pagrindimas remiantis kintamuoju3 (Užimti gyventojai (tūkst. mot.))

Taigi tiek suradę pagrindines skaitines padėties charakteristikas, tiek išbrėžią linijinį grafiką, matome, jog didžiausia kintamojo3 reikšmė buvo 1571 tūkst. (2017 m.). O štai minimali kintamojo3 reikšmė buvo 1401 (2004 m.). Sunku įvardinti, kas galėjo nulemti tokį moterų užimtumą Austrijoje.

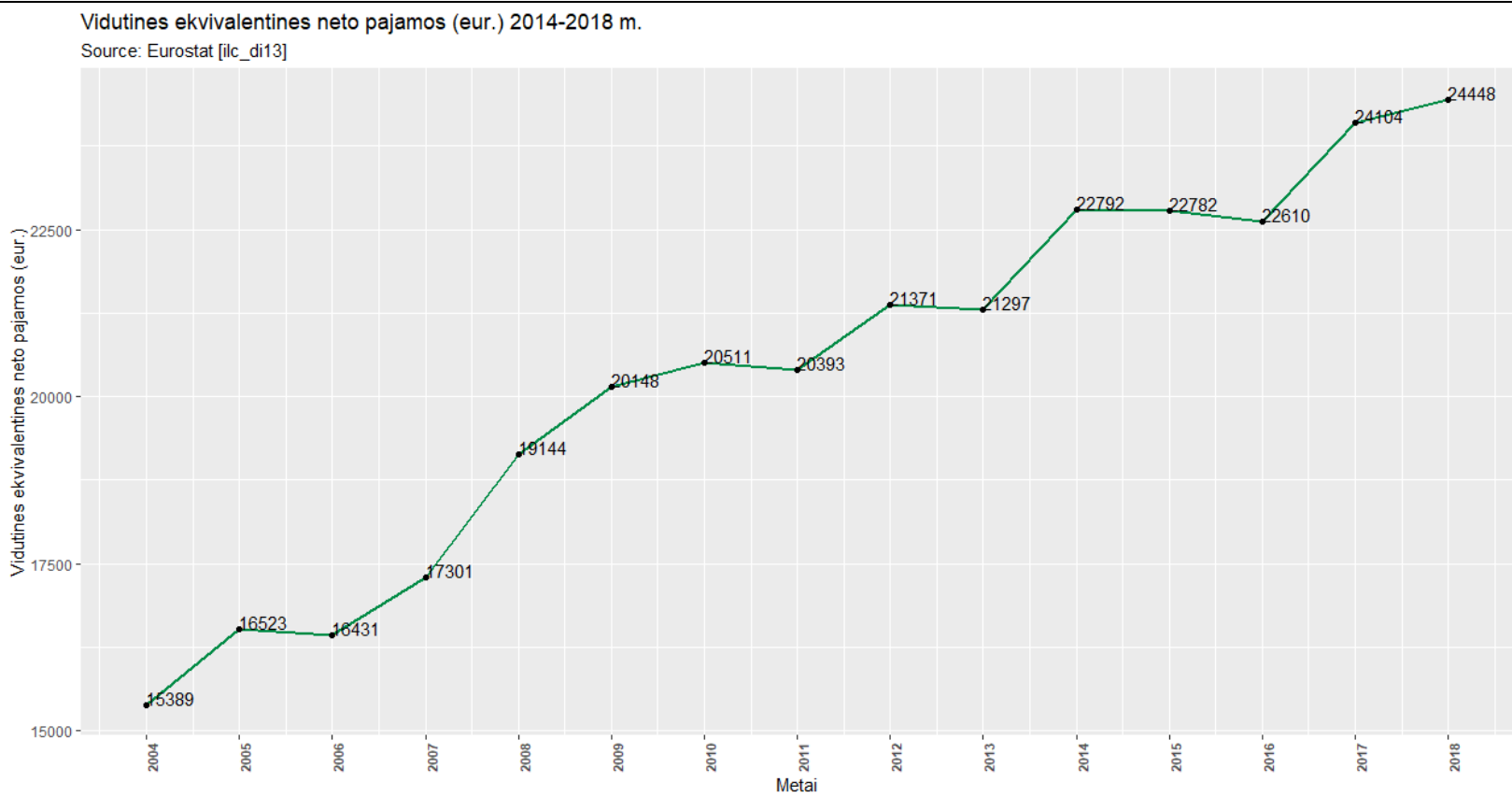
Taip pat iš linijinio grafiko matome, jog moterų užimtumas Austrijoje turi tendenciją didėti (išskyrus 2018 m., nes tais metais užimtumas buvo šiek tiek mažesnis nei 2017 m.). Moterų užimtumo didėjimą Austrijoje būtų galima sieti su ekonomikos augimu (augant ekonomikai, yra sukuriama daugiau darbo vietų, mažinamas nedarbo lygis šalyje. Todėl galbūt auganti ekonomika ir turėjo didelę reikšmę moterų užimtumo didėjimui). Dauguma žmonių mano, jog recesijos laikotarpiu (šiuo atveju 2008-2009m.) nedarbo lygis turi ženkliai išaugti, tačiau tikriausiai pamirštama tai, jog valstybė norėdama išgelbėti savo šalį nuo recesijos renkasi tokią politiką, kuri skatintų vartojimą (vartojimas skatina ekonomikos augimą, o ekonomikos augimas padeda atsigauti po recesijos). Vartojimas gali būti paskatintas per vyriausybės transferus (pašalpos bedarbiams, stipendijos studentams, pensijos pensininkams, pašalpos daugiavaikėms šeimoms ir kt. socialinės išmokos) ir užtikrinant, jog individai turės darbo vietą (žiūrint į grafiką, galima manyti, jog recesijos metu Austrijos vyriausybė stengėsi užtikrinti, jog užimtumas nemažėtų). Taip pat augantis moterų užimtumas gali būti siejamas ir su įvairiais judėjimais, kurių tikslas – siekti lygių teisių su vyrais.

```
ggplot(data=macro, aes(x=Laikas, y=kintamasis4, label=kintamasis4))+
  geom_line(color="springgreen4", size=1)+
  geom_point()+
```

```

geom_text(aes(label=kintamasis4), hjust=0, vjust=0)+
scale_x_continuous(breaks=macro$Laikas, labels = macro$Laikas)+
theme(axis.text.x = element_text(face="bold", size=8,
angle=90))+
labs(x="Metai", y="Vidutinės ekvivalentinės neto pajamos
(eur.)", subtitle = "Source: Eurostat [ilc_di13]")+
ggtitle("Vidutinės ekvivalentinės neto pajamos (eur.) 2014-
2018 m.")

```



### **Ekonominis pagrindimas remiantis kintamuoju4 (Vidutinės ekvivalentinės neto pajamos (eur.))**

Taigi tiek suradę pagrindines skaitines padėties charakteristikas, tiek išbrėžią linijinį grafiką, matome, jog didžiausia kintamojo4 reikšmė buvo 24448 e. (2018 m.), taip yra todėl, jog 2018 metais tiek pasaulinė, tiek ES (ir tuo pačiu Austrijos) ekonomika buvo susidūrusi su labai sparčiu ekonomikos augimu. Įprasta, jog augant ekonomikai, auga ir vidutinės ekvivalentinės neto pajamos.

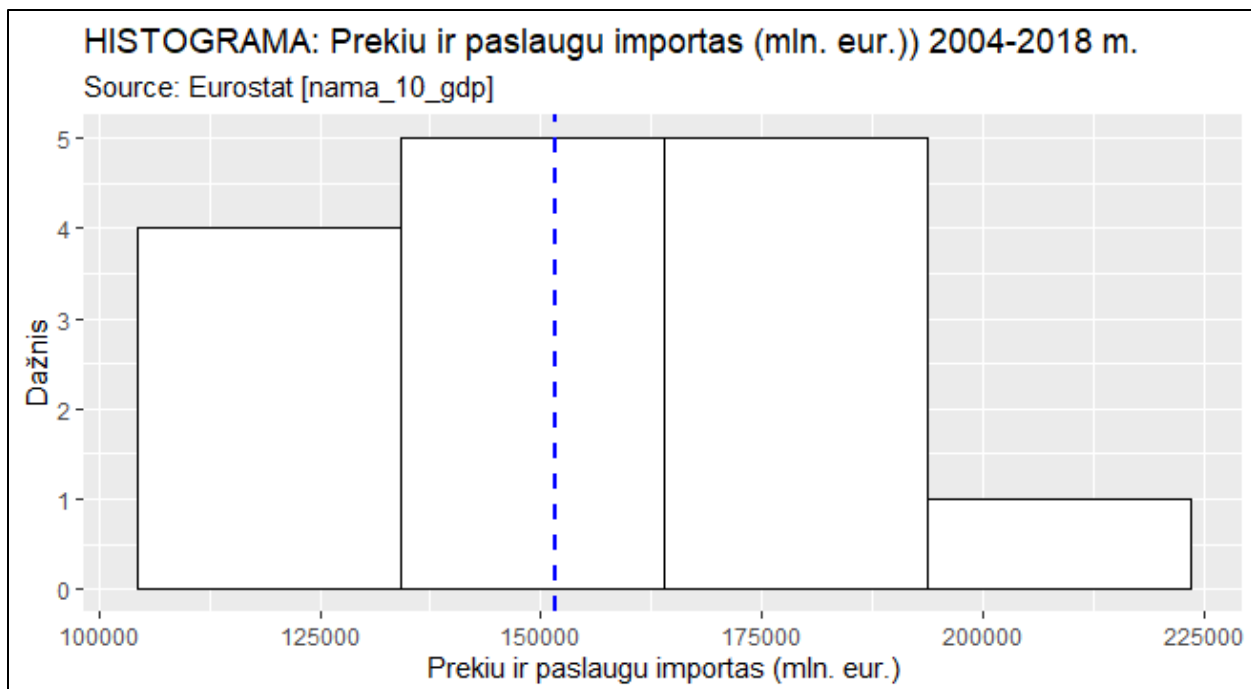
Minimali kintamojo4 reikšmė buvo 15389 e. (2004 m.), sunku įvardinti pagrindinius veiksnius, kurie tai nulėmė, nes 2004 m. Austrijoje ekonominė situacija buvo labai gera (2004 m. Austrija buvo ketvirta pagal turtingumą šalis ES, ją pralenkė tik Liuksemburgas, Airija ir Nyderlandai.). Mano manymu, vidutinės ekvivalentinės neto pajamos 2004 m. buvo pakankamai

žemos, nes tuo metu ES ekonomika nebuvo pasiekusi tokio atvirumo lygio kaip dabar (t.y. šalių ekonomikos buvo kur kas uždaresnės, mažiau tarptautinių kompanijų buvo įsiliejusios į šalies vidų). Kaip žinome iš makroekonomikos teorijos, kuo daugiau rinkos dalyvių yra rinkoje (šiuo atveju kuo daugiau kompanijų yra šalyje) tuo didesnė tarp jų vyrauja konkurenciją (nenordamos pralaimėti konkurencinėje kovoje, jos įvairiais būdais stengiasi pritraukti reikiamą darbuotojų skaičių, šiuo atveju galima daryti prielaidą, jog užsienio šalių kompanijos įkūrusios savo padalinius ir Austrijoje, stengėsi didinti darbo užmokestį, jog sudomintų kuo daugiau individų, kurie nori ten dirbti. Tai paaiškintų kodėl bėgant laikui Austrijos vidutinės ekvivalentinės neto pajamos augo (žinoma, buvo laikotarpį, kai vidutinės ekvivalentinės neto pajamos sumažėdavo lyginant su praėjusiais metais). Taip yra todėl, jog vidutinėms ekvivalentinėms neto pajamos įtakos turi ne tik ekonomikos augimas, bet ir vyriausybės politiniai sprendimai.

```
#Kad nustatyčiau bandwidth optimalų dydį naudodsiu šią taisyklę
h=2*IQR*n-1/3
bw <- 2 * IQR(macro$kintamasis1) / length(macro$kintamasis1)^(1/3)

ggplot(data=macro, aes(x=kintamasis1)) +
  geom_histogram(binwidth=bw, color="black", fill="white")+
  ggtitle("HISTOGRAMA: Prekių ir paslaugų importas (mln. eur.)
2004-2018 m.")+
  labs(y="Dažnis", x="Prekių ir paslaugų importas (mln. eur.)",
  subtitle ="Source: Eurostat [nama_10_gdp]")+
  geom_vline(aes(xintercept=mean(kintamasis1)),
             color="blue", linetype="dashed", size=1)

#Kad sužinočiau, kam lygus vidurkis, jį randu taip:
mean(macro$kintamasis1)
[1] 151685.7
#Mėlyna linija žymi vidurkį, kuris yra lygus 151685.7
```

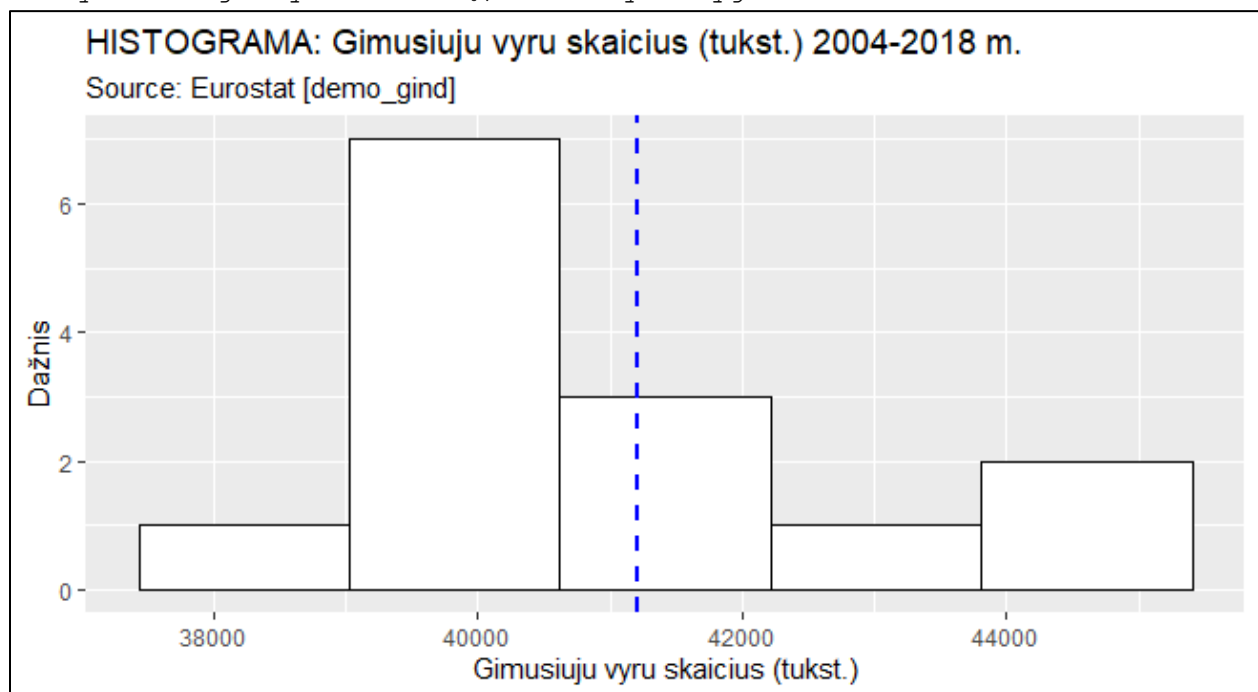


```

#Kadangi kintamsis2 2018 m. turi NA reikšmę, tai sukuriu naują df2,
kuriame nebus 2018 metų
library(dplyr)
df2 <- macro %>%
  filter(Laikas <= "2017")

bw1 <- 2 * IQR(df2$kintamasis2) / length(df2$kintamasis2)^(1/3)
ggplot(data=df2, aes(x=kintamasis2)) +
  geom_histogram(binwidth=bw1, color="black", fill="white")+
  ggtitle("HISTOGRAMA: Gimusiųjų vyrų skaičius (tūkst.) 2004-
2018 m.")+
  labs(y="Dažnis", x="Gimusiųjų vyrų skaičius (tūkst.)",
  subtitle ="Source: Eurostat [demo_gind]")+
  geom_vline(aes(xintercept=mean(df2$kintamasis2)),
    color="blue", linetype="dashed", size=1)
#Kad sužinočiau, kam lygus vidurkis, jį randu taip:
mean(df2$kintamasis2)
[1] 41205.14
#Mėlyna linija žymi vidurkį, kuris yra lygus 41205.14

```



```

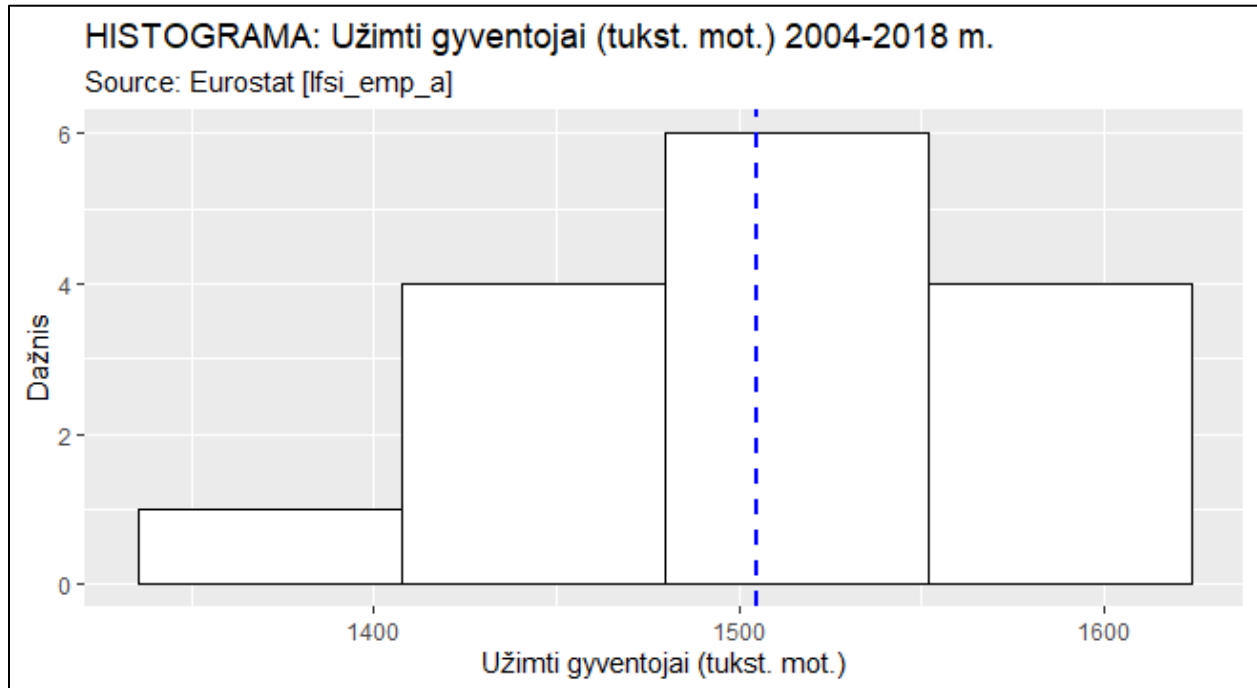
bw2 <- 2 * IQR(macro$kintamasis3) / length(macro$kintamasis3)^(1/3)
ggplot(data=macro, aes(x=kintamasis3)) +
  geom_histogram(binwidth=bw2, color="black", fill="white")+
  ggtitle("HISTOGRAMA: Užimti gyventojai (tūkst. mot.) 2004-2018
m.")+
  labs(y="Dažnis", x="Užimti gyventojai (tūkst. mot.)", subtitle
="Source: Eurostat [lfsi_emp_a]")+
  geom_vline(aes(xintercept=mean(kintamasis3)),

```

```

        color="blue", linetype="dashed", size=1)
#Kad sužinočiau, kam lygus vidurkis, jį randu taip:
mean(macro$kintamasis3)
[1] 1504.933
#Mėlyna linija žymi vidurkį, kuris yra lygus 1504.933

```

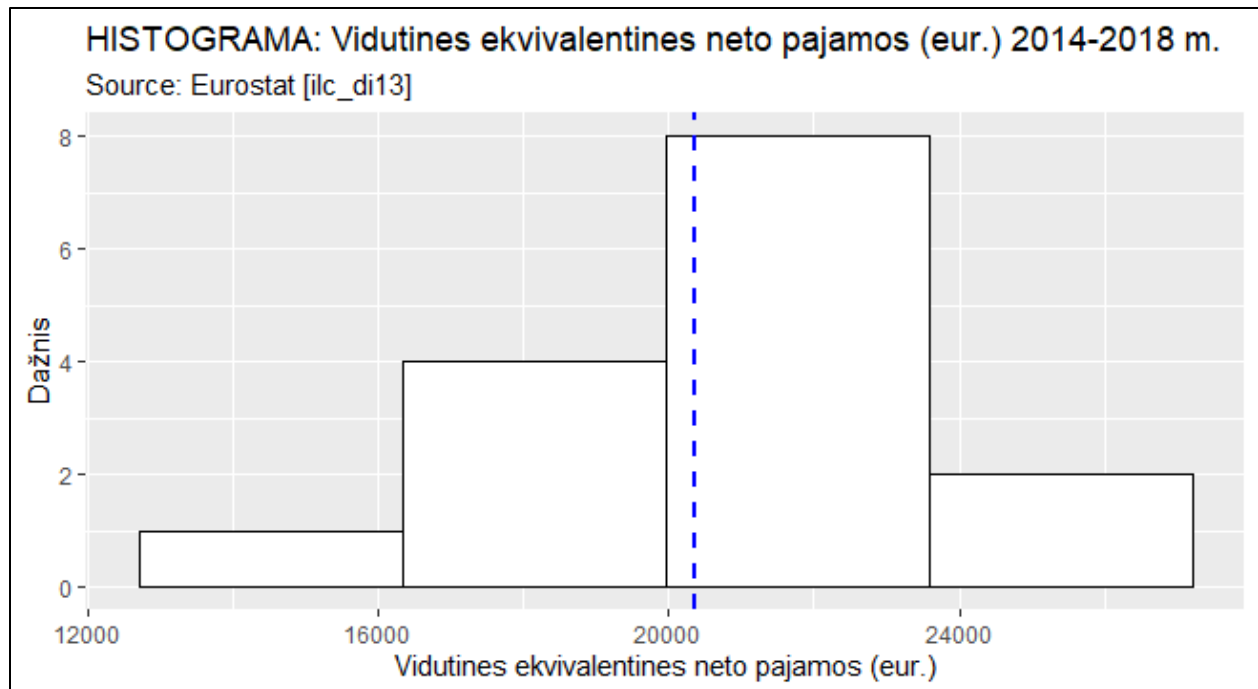


```

bw3 <- 2 * IQR(macro$kintamasis4) / length(macro$kintamasis4)^(1/3)
ggplot(data=macro, aes(x=kintamasis4)) +
  geom_histogram(binwidth=bw3, color="black", fill="white")+
  ggtitle("HISTOGRAMA: Vidutinės ekvivalentinės neto pajamos
(eur.) 2014-2018 m.")+
  labs(y="Dažnis", x="Vidutinės ekvivalentinės neto pajamos
(eur.)", subtitle = "Source: Eurostat [ilc_di13]")+
  geom_vline(aes(xintercept=mean(kintamasis4)),
    color="blue", linetype="dashed", size=1)

#Kad sužinočiau, kam lygus vidurkis, jį randu taip:
mean(macro$kintamasis4)
[1] 20349.6
#Mėlyna linija žymi vidurkį, kuris yra lygus 20349.6

```



5. Rinkmenoje „apklausa“ pasirinkite du kintamuosius ir atlikite t-testą. (Nepamirškite parašyti tyrimo hipotezę, statistinę hipotezę bei tyrimo išvadą).

**Tyrimo hipotezė:** Moterys uždirba tiek pat kiek ir vyrai iš samdomojo darbo pajamų (per metus)

**Statistinė hipotezė**  $\begin{cases} H_0: \mu_x = \mu_y \\ H_1: \mu_x \neq \mu_y \end{cases}$  čia, x- moterų samdomojo darbo užmokestis (per metus),

y – vyrų samdomojo darbo užmokestis (per metus)

```
#PY010G - Samdomojo darbo pajamos(per metus), RB090 - lytis
t.test(apklausa$PY010G~as.factor(apklausa$RB090))
```

**welch Two sample t-test**

```
data: apklausa$PY010G by as.factor(apklausa$RB090)
t = 3.3771, df = 10344, p-value = 0.0007353
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 377.8166 1423.2083
sample estimates:
mean in group Vyras mean in group Moteris
 9036.664           8136.151
```

**Tyrimo išvada** Kadangi p reikšmė (angl. p-value) yra mažesnė už 0.05,  $H_0$  atmetame ir tariame, kad teisinga hipotezė  $H_1$  (t.y. moterys uždirba ne tiek pat kiek vyrai iš samdomojo darbo pajamų per metus)

6. Rinkmenoje „apklausa“ pasirinkite intervalinį kintamąjį bei suskaičiuokite vidurkio ir dispersijos taškinius įverčius bei vidurkio pasikliautinąjį intervalą.



```

rez<-summarise(apklausa,vid=mean(PY010G, na.rm=T), std=sd(PY010G,
na.rm=T),disp=var(PY010G, na.rm=T)) # papildomai standartinis
nuokrypis
rez
# A tibble: 1 x 3
  vid      std      disp
  <dbl>  <dbl>  <dbl>
1 8548. 13854. 191941585.
error <- qt(0.975,df=length(apklausa$PY010G-1))*sd(apklausa$PY010G,
na.rm=T)/sqrt(length(apklausa$PY010G)) # pasikliovimo lygmuo 0,95
error
[1] 242.9734
left <- mean(apklausa$PY010G, na.rm=T)-error
right <- mean(apklausa$PY010G, na.rm=T)+error
left
[1] 8304.81
right
[1] 8790.757

```

---

7. 6-oje dalyje gautus taškinius įverčius panaudokite atsitiktinių normalių dydžių generavimui. Porai (vidurkis, dispersija) sugeneruokite 100 normalių atsitiktinių dydžių imčių (imties dydis 500 stebėjimų). Apskaičiuokite gautų imčių taškinius vidurkio bei dispersijos įverčius. Gautiems vidurkių taškiniams įverčiams išbrėžkite histogramą bei stačiakampę diagramą. Ką galite pasakyti apie įverčių stabilumą?

```

observations <- matrix(rnorm(50000, mean=rez$vid, sd=rez$std), 100, 50
0) #čia naudojamos anksčiau apskaičiuotos charakteristikos
means<- apply(observations,1,mean)
variance<-apply(observations,1,var)
deviation<-apply(observations,1,sd) # papildomai std nuokrypis

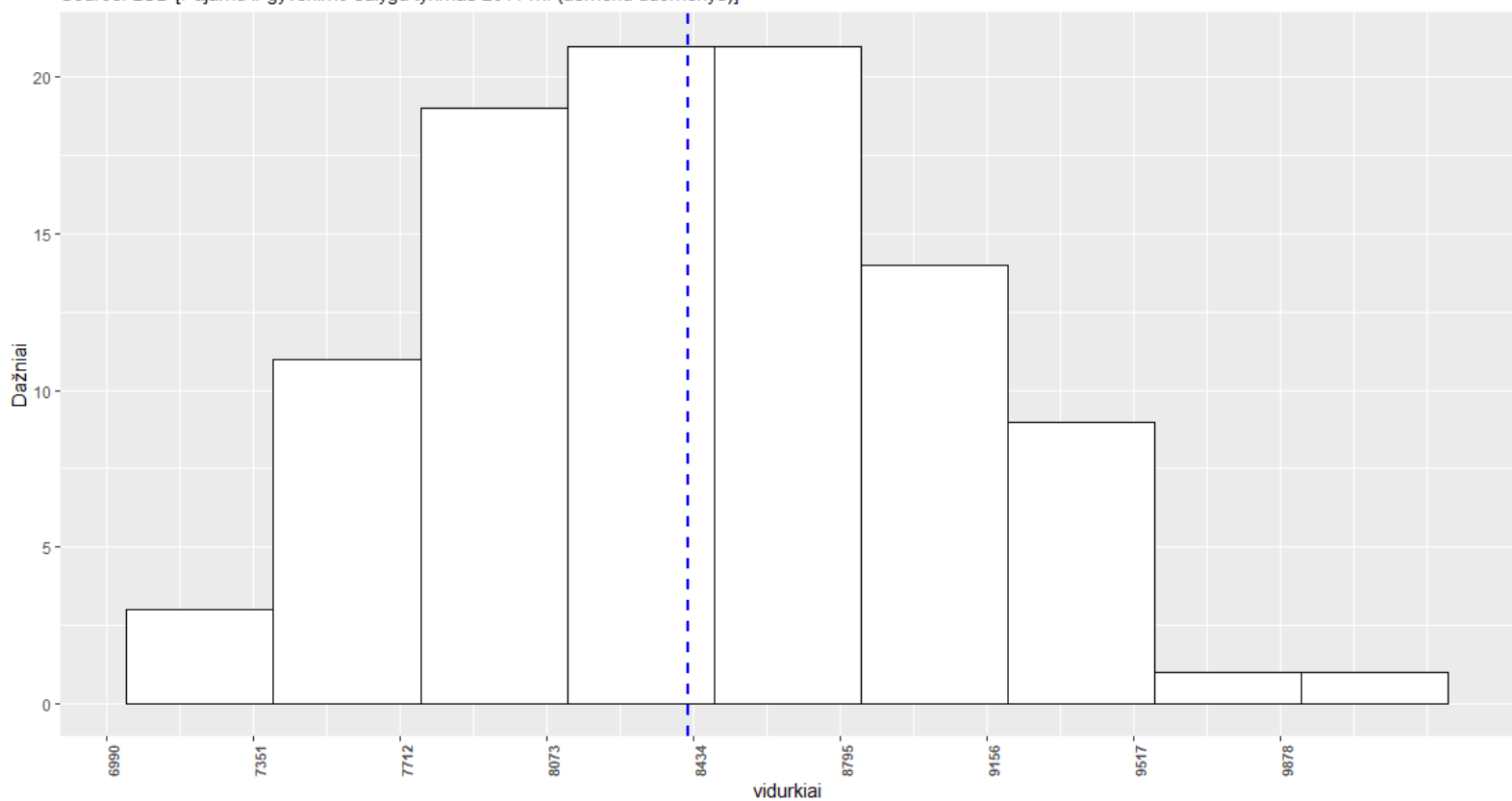
bw3 <- round(2 * IQR(means) / length(means)^(1/3))
ggplot(as.data.frame(means),aes(x=means))+
  geom_histogram(binwidth=bw3,color="black", fill="white")+
  scale_y_continuous(breaks=seq(0,30,5), labels=seq(0,30,5))+
  scale_x_continuous(breaks=seq(6990,10119, bw3),labels = seq(69
90,10119, bw3))+
  theme(axis.text.x = element_text(face="bold", size=8, angle=9
0))+
  labs(x="vidurkiai",y="Dažniai", subtitle="Source: LSD [Pajamų
ir gyvenimo sąlygų tyrimas 2011 m. (asmenų duomenys)]")+
  ggtitle("Vidurkių įverčių histograma")+
  geom_vline(aes(xintercept=mean(means)),
             color="blue", linetype="dashed", size=1)

mean(means)
[1] 8418.768
#Mėlyna linija žymi vidurkį, kuris yra lygus 8418.768

```

## Vidurkiu iverciu histograma

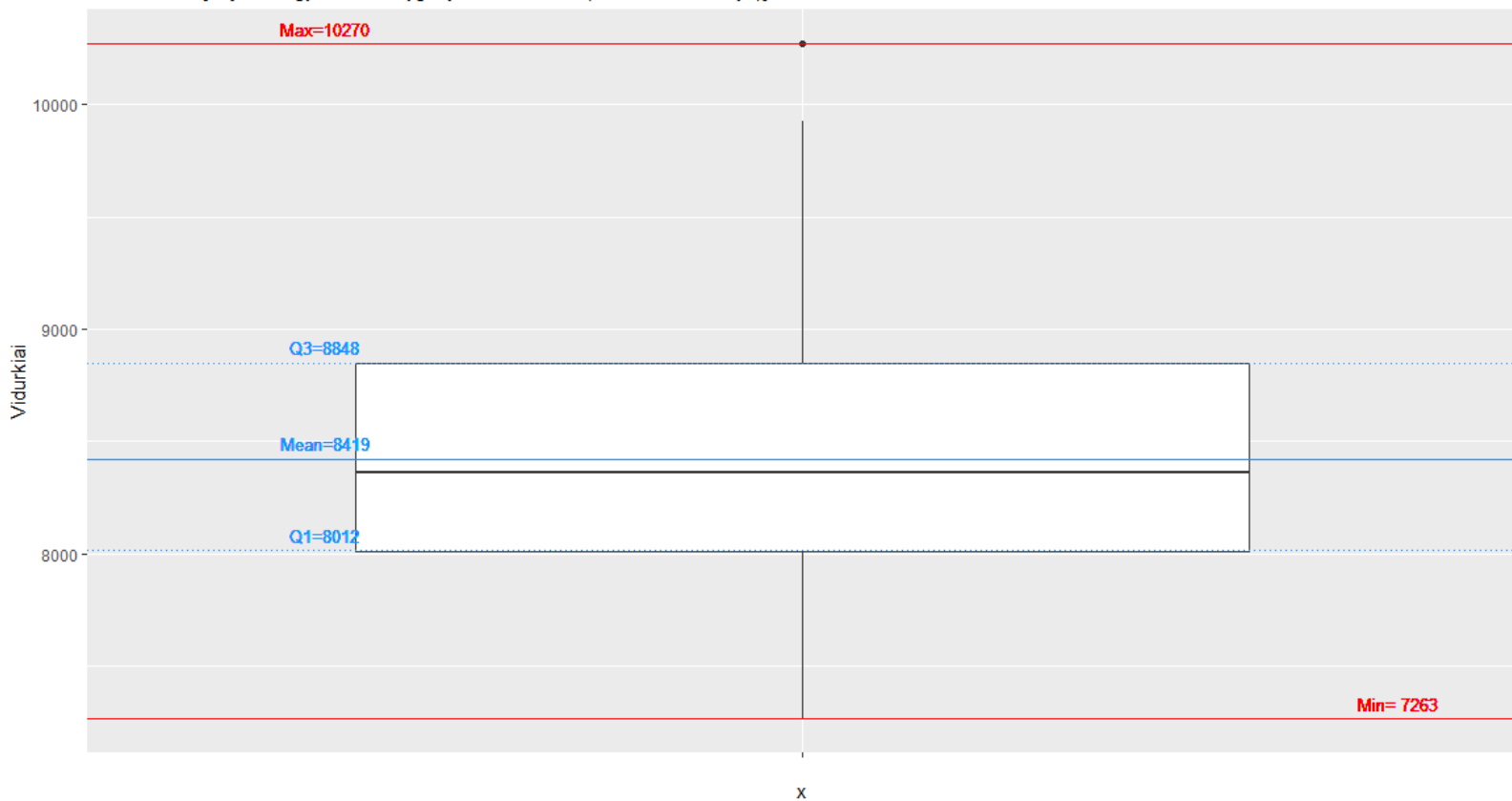
Source: LSD [Pajamų ir gyvenimo sąlygų tyrimas 2011 m. (asmenų duomenys)]



```
#panaduojuosi summary f-ja randu reikimas charakteristikas
summary(means)
Min. 1st Qu. Median Mean 3rd Qu. Max.
 7263   8012   8362   8419   8848  10270
ggplot(as.data.frame(means), aes(x="", y=means)) +
  geom_boxplot() +
  ggtitle("Stačiakampė diagrama") +
  labs(y="Vidurkiai", subtitle="Source: LSD [Pajamų ir gyvenimo
sąlygų tyrimas 2011 m. (asmenų duomenys)]") +
  geom_hline(yintercept=8419, linetype="solid", color = "dodgerb
lue") +
  geom_text(size=3.5, aes(0.6, 8419, label = "Mean=8419", vjust =
-0.6), colour="dodgerblue") +
  geom_hline(yintercept=8848, linetype="dotted", color = "dodger
blue") +
  geom_text(size=3.5, aes(0.6, 8848, label = "Q3=8848", vjust = -0
.6), colour="dodgerblue") +
  geom_hline(yintercept=10270, linetype="solid", color = "red") +
  geom_text(size=3.5, aes(0.6, 10270, label = "Max=10270", vjust =
-0.6), colour="red") +
  geom_hline(yintercept=8012, linetype="dotted", color = "dodger
blue") +
  geom_text(size=3.5, aes(0.6, 8012, label = "Q1=8012", vjust = -0
.6), colour="dodgerblue") +
  geom_hline(yintercept= 7263, linetype="solid", color = "red") +
  geom_text(size=3.5, aes(1.5, 7263, label = "Min= 7263", vjust =
-0.6), colour="red")
```

## Staciakampe diagrama

Source: LSD [Pajamu ir gyvenimo salygu tyrimas 2011 m. (asmenu duomenys)]



Mano nuomone, įverčiai yra gana stabilūs, nes yra tik viena išskirtis (tai matosi iš stačiakampės diagramos). Taip pat įverčių reikšmės svyruoja nuo 7263 iki 10270 (o tai mano nuomone, yra sąlyginai mažas skirtumas)

8. Rinkmenos „macro“ kintamiesiems išskirkite tiesinius trendus. Pavaizduokite duomenų ir trendų grafikus. Apibūdinkite laiko eilučių tendencijas.

```
library(dplyr)
```

```
library(car)
```

```
trnd<-lm(kintamasis1~Laikas, data=macro, x=T, y=T)
```

```
trnd
```

```
Call:
```

```
lm(formula = kintamasis1 ~ Laikas, data = macro, x = T, y = T)
```

```
Coefficients:
```

```
(Intercept)      Laikas  
-11641197      5864
```

```
summary(trnd)
```

```
Call:
```

```
lm(formula = kintamasis1 ~ Laikas, data = macro, x = T, y = T)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max  
-19397.1 -4346.5   565.8   6244.9   9335.5
```

```
Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -1.164e+07  9.566e+05 -12.17 1.77e-08 ***  
Laikas       5.864e+03  4.757e+02  12.33 1.51e-08 ***
```

```
---
```

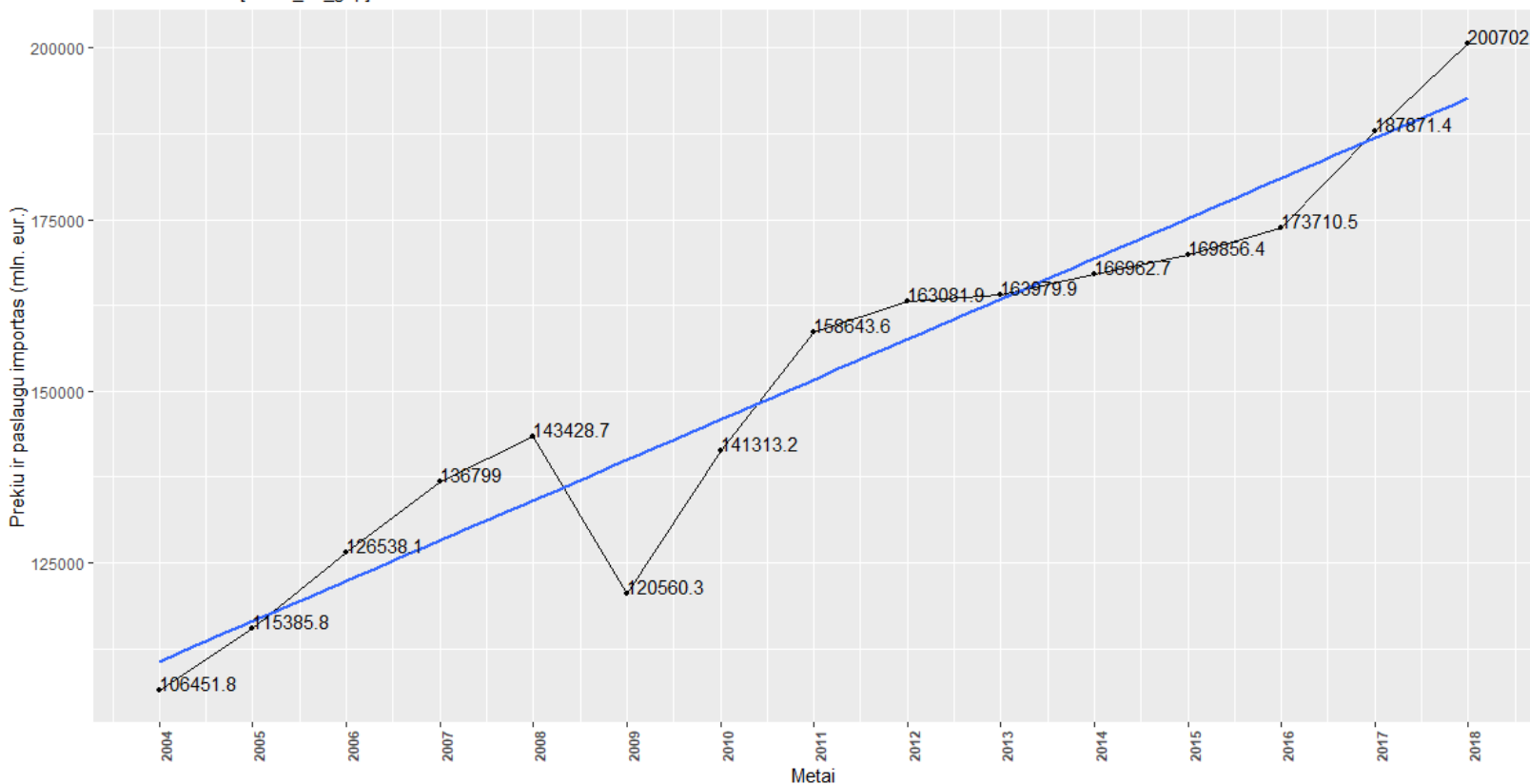
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 7960 on 13 degrees of freedom  
Multiple R-squared: 0.9212, Adjusted R-squared: 0.9151  
F-statistic: 152 on 1 and 13 DF, p-value: 1.512e-08

```
# grafikas
library(ggplot2)
ggplot(macro, aes(Laikas, kintamasis1))+
  scale_x_continuous(breaks=macro$Laikas, labels=macro$Laikas)+
  geom_line(stat="identity")+
  geom_point(size=1)+
  stat_smooth(method = "lm", se = FALSE)+
  geom_text(aes(label=kintamasis1), hjust=0, vjust=0)+
  theme(axis.text.x = element_text(face="bold", size=8,
angle=90))+
  labs(x="Metai", y="Prekių ir paslaugų importas (mln. eur.)",
  subtitle = "Source: Eurostat [nama_10_gdp]")+
  ggtitle("Prekių ir paslaugų importas (mln. eur.) 2004-2018
m.")
```

Prekių ir paslaugų importas (mln. eur.) 2004-2018 m.

Source: Eurostat [nama\_10\_gdp]



Trendo linija rodo prekių ir paslaugų importo matuojamo mln. eur. bendrąjį padidėjimą.

```

trnd<-lm(kintamasis2~Laikas, data=macro, x=T, y=T)
trnd
Call:
lm(formula = kintamasis2 ~ Laikas, data = macro, x = T, y = T)

Coefficients:
(Intercept)      Laikas
-775835.5      406.4
summary(trnd)

Call:
lm(formula = kintamasis2 ~ Laikas, data = macro, x = T, y = T)

Residuals:
    Min       1Q   Median       3Q      Max
-1523.6  -955.2  -264.3   823.5  1976.4

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -775835.55   159055.90  -4.878  0.000380 ***
Laikas        406.39      79.11    5.137  0.000246 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

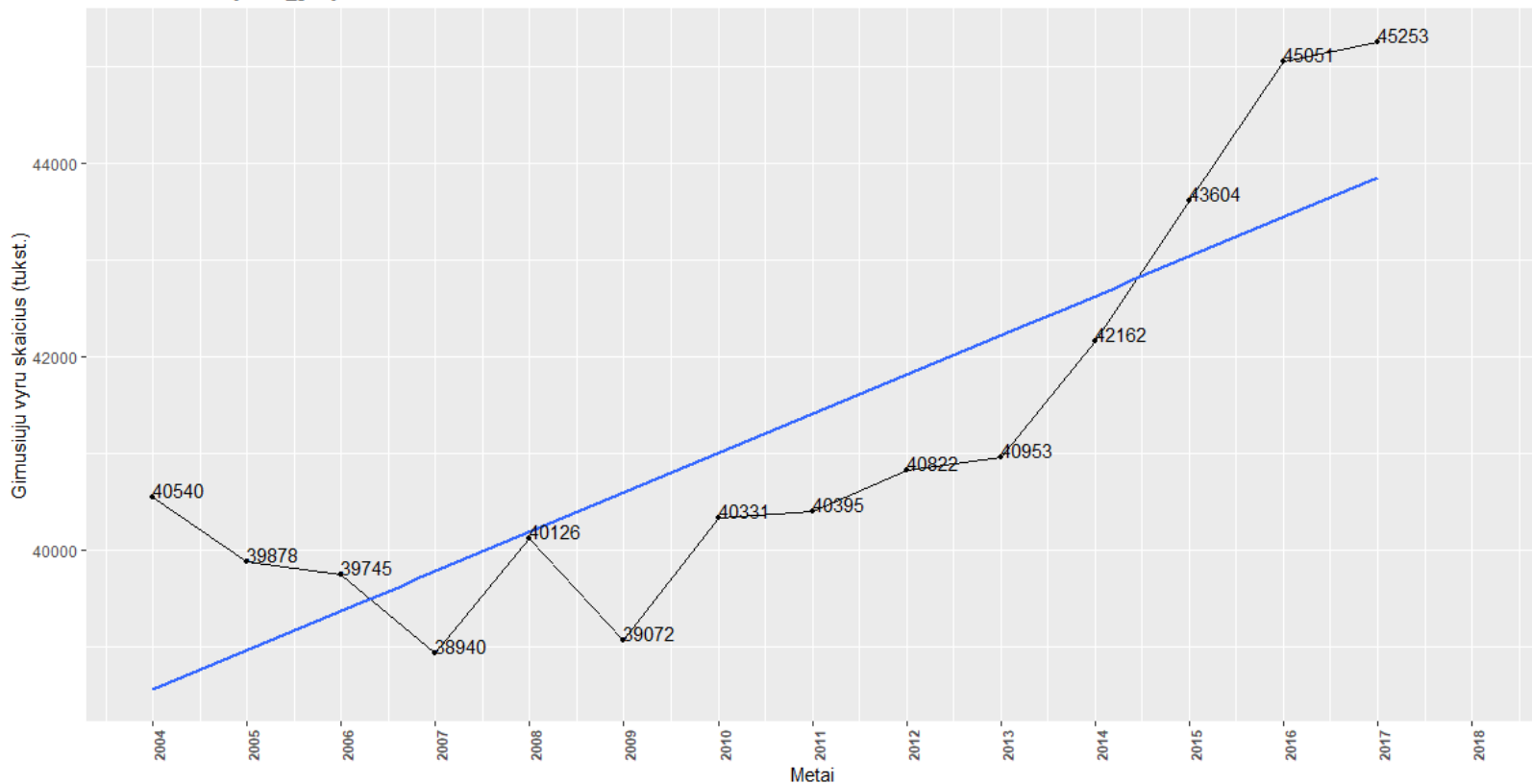
Residual standard error: 1193 on 12 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.6874, Adjusted R-squared:  0.6613
F-statistic: 26.39 on 1 and 12 DF, p-value: 0.0002463

ggplot(macro, aes(Laikas, kintamasis2))+
  scale_x_continuous(breaks=macro$Laikas,labels=macro$Laikas)+
  geom_line(stat="identity")+
  geom_point(size=1)+
  stat_smooth(method = "lm", se = FALSE)+
  geom_text(aes(label=kintamasis2), hjust=0, vjust=0)+
  theme(axis.text.x = element_text(face="bold", size=8, angle=90))+
  labs(x="Metai", y="Gimusiųjų vyrų skaičius (tūkst.)", subtitle
="Source: Eurostat [demo_gind]")+
  ggtitle("Gimusiųjų vyrų skaičius (tūkst.) 2004-2018 m.")

```

# Gimusiųjų vyrų skaičius (tūkst.) 2004-2018 m.

Source: Eurostat [demo\_gind]



Trendo linija rodo gimusiųjų vyrų skaičiaus matuojamo tūkstančiais bendrąjį padidėjimą.

```
trnd<-lm(kintamasis3~Laikas, data=macro, x=T, y=T)
```

```
trnd
```

```
Call:
```

```
lm(formula = kintamasis3 ~ Laikas, data = macro, x = T, y = T)
```

```
Coefficients:
```

```
(Intercept)      Laikas  
-23302.19       12.34
```

```
summary(trnd)
```

```
Call:
```

```
lm(formula = kintamasis3 ~ Laikas, data = macro, x = T, y = T)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max  
-28.283  -5.601   1.074   9.735  14.395
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) -2.330e+04  1.547e+03  -15.06 1.31e-09 ***  
Laikas       1.234e+01  7.691e-01   16.04 6.03e-10 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

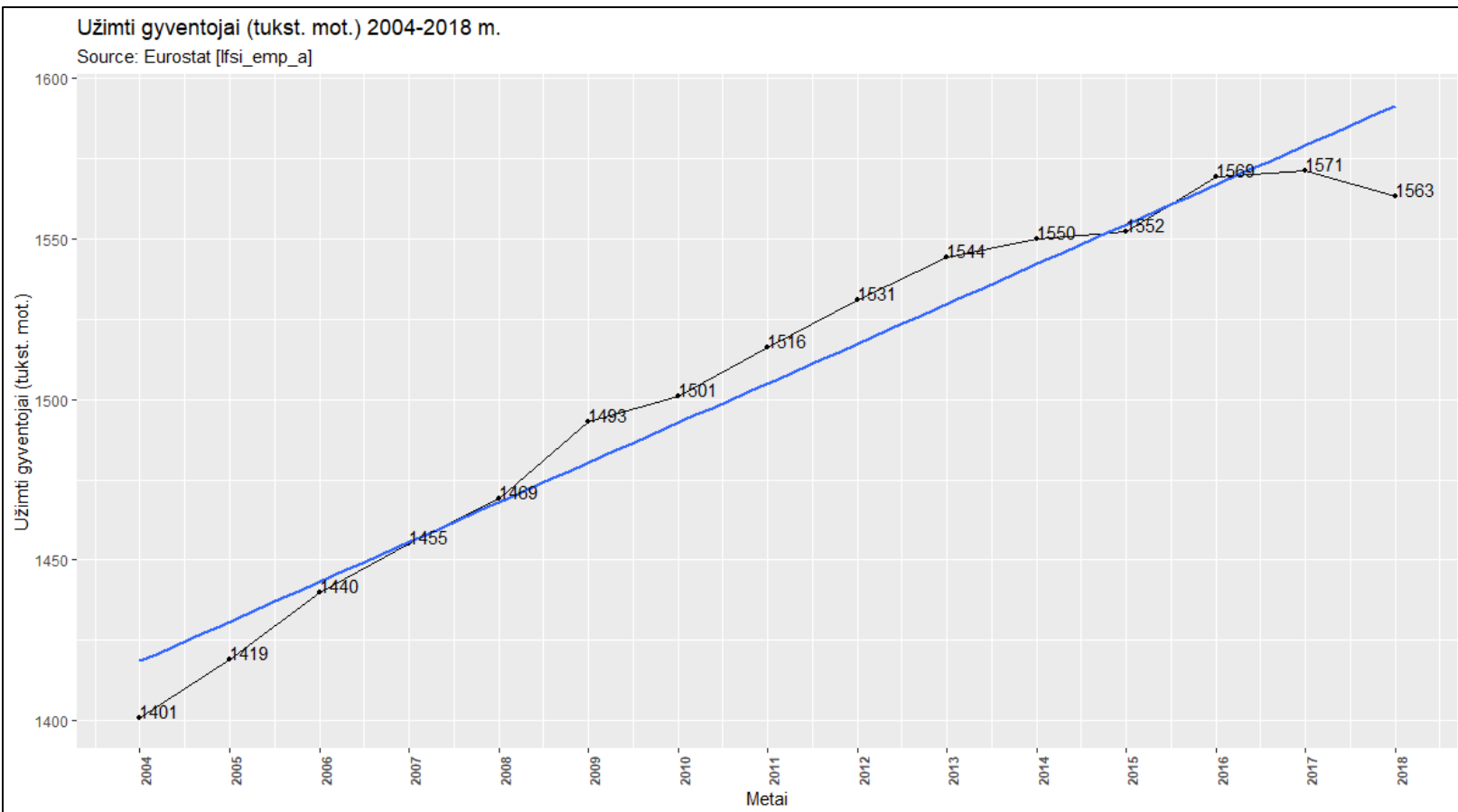
```
Residual standard error: 12.87 on 13 degrees of freedom
```

```
Multiple R-squared:  0.9519, Adjusted R-squared:  0.9482
```

```
F-statistic: 257.2 on 1 and 13 DF, p-value: 6.03e-10
```

```
ggplot(macro, aes(Laikas, kintamasis3))+  
  scale_x_continuous(breaks=macro$Laikas, labels=macro$Laikas)+  
  geom_line(stat="identity")+  
  theme_minimal()
```

```
geom_point(size=1)+
stat_smooth(method = "lm", se = FALSE)+
geom_text(aes(label=kintamasis3), hjust=0, vjust=0)+
theme(axis.text.x = element_text(face="bold", size=8, angle=90))+
labs(x="Metai", y="Užimti gyventojai(tūkst. Mot.)", subtitle = "Source: Eurostat [lfsi_emp_a]")+
ggtitle("Užimti gyventojai(tūkst. Mot.)2004-2018 m.")
```



Trendo linija rodo moterų užimtumo matuojamo tūkst. bendrąjį padidėjimą.

```
trnd<-lm(kintamasis4~Laikas, data=macro, x=T, y=T)
trnd
Call:
lm(formula = kintamasis4 ~ Laikas, data = macro, x = T, y = T)

Coefficients:
(Intercept)      Laikas
-1242414.8       627.9

summary(trnd)
Call:
lm(formula = kintamasis4 ~ Laikas, data = macro, x = T, y = T)

Residuals:
    Min       1Q   Median       3Q      Max
-879.24 -422.67  -59.03   476.04 1054.26
```

Coefficients:

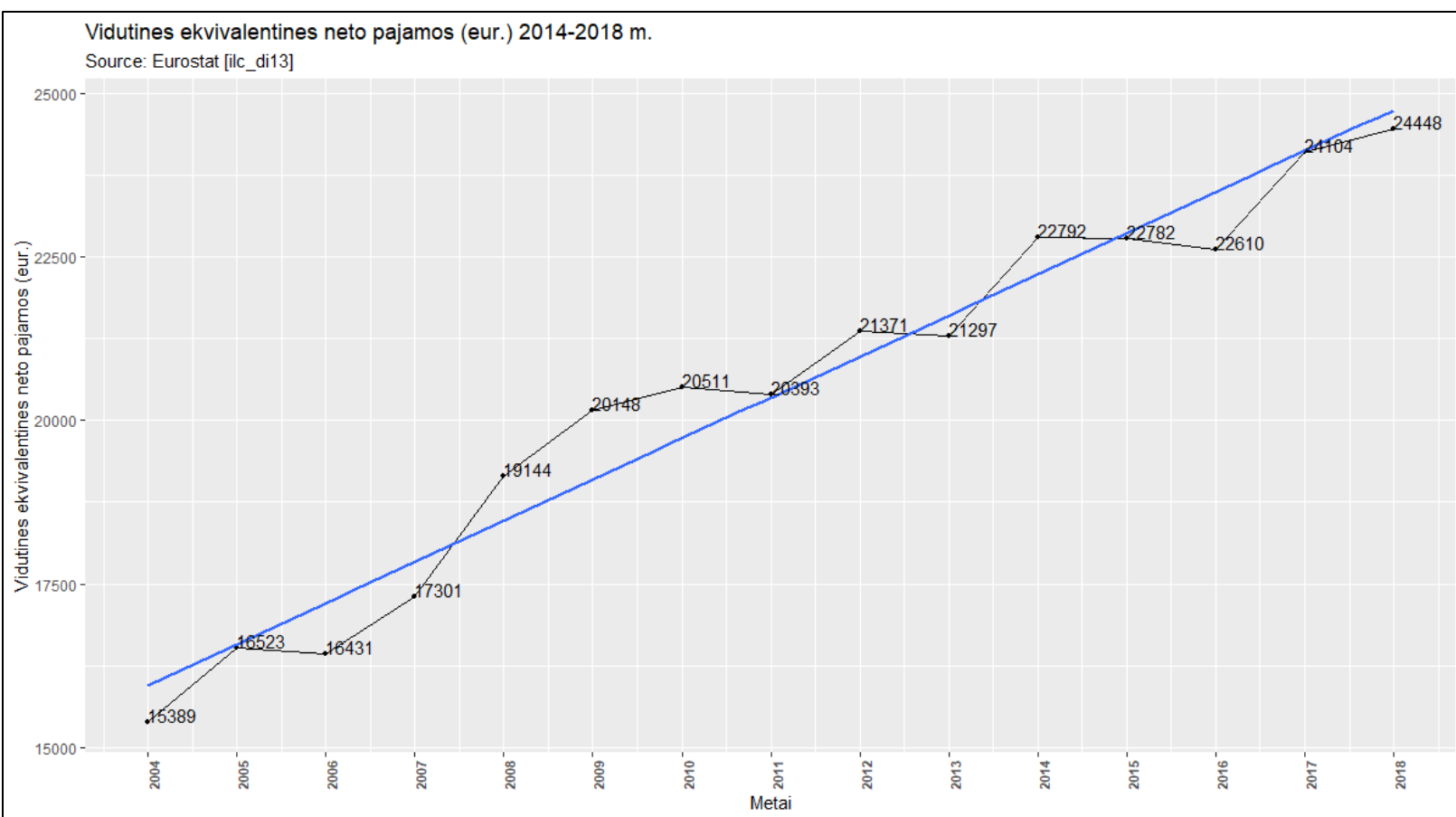
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-1.242e+06	7.336e+04	-16.94	3.06e-10	***
Laikas	6.279e+02	3.648e+01	17.21	2.50e-10	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 610.4 on 13 degrees of freedom  
Multiple R-squared: 0.958, Adjusted R-squared: 0.9547  
F-statistic: 296.3 on 1 and 13 DF, p-value: 2.5e-10

```
ggplot(macro, aes(Laikas, kintamasis4))+  
  scale_x_continuous(breaks=macro$Laikas, labels=macro$Laikas)+  
  geom_line(stat="identity")+  
  geom_point(size=1)+  
  stat_smooth(method = "lm", se = FALSE)+  
  geom_text(aes(label=kintamasis4), hjust=0, vjust=0)+  
  theme(axis.text.x = element_text(face="bold", size=8, angle=9  
0))+  
  labs(x="Metai", y="Vidutinės ekvivalentinės neto pajamos (eur.  
)", subtitle="Source: Eurostat [ilc_di13]")+  
  ggtitle("Vidutinės ekvivalentinės neto pajamos (eur.) 2014-201  
8 m.")
```



Trendo linija rodo vidutinių ekvivalentinių neto pajamų matuojamų eurai bendrąjį padidėjimą.



---

9. Pasirinkite kelis arba visus „macro“ duomenų kintamuosius ir sudarykite tiesinę regresiją ir atlikite prognozę. Interpretuokite gautus rezultatus

```
#Kadangi kintamasis2 turi NA reikšmę, todėl pasirenku panaudoti  
use="complete.obs"  
cor(macro[-1], use="complete.obs") #apskaičiuoja koreliacijas visiems  
kintamiesiems, kintamojo su savimi pačiu koreliacija lygi 1
```

```
      kintamasis1 kintamasis2 kintamasis3 kintamasis4  
kintamasis1  1.0000000  0.7762114  0.9387647  0.9118741  
kintamasis2  0.7762114  1.0000000  0.7333548  0.7405589  
kintamasis3  0.9387647  0.7333548  1.0000000  0.9801467  
kintamasis4  0.9118741  0.7405589  0.9801467  1.0000000
```

```
cr <- corr.test(macro[,-1], ci=TRUE, use="complete.obs") #tikrina ar k  
iekviena koreliacija statistiskai reikšminga
```

```
cr
```

```
Call:corr.test(x = macro[, -1], use = "complete.obs", ci = TRUE)
```

```
Correlation matrix
```

```
      kintamasis1 kintamasis2 kintamasis3 kintamasis4  
kintamasis1  1.00      0.78      0.94      0.91  
kintamasis2  0.78      1.00      0.73      0.74  
kintamasis3  0.94      0.73      1.00      0.98  
kintamasis4  0.91      0.74      0.98      1.00
```

```
Sample Size
```

```
      kintamasis1 kintamasis2 kintamasis3 kintamasis4  
kintamasis1    15      14      15      15  
kintamasis2    14      14      14      14  
kintamasis3    15      14      15      15  
kintamasis4    15      14      15      15
```

```
Probability values (Entries above the diagonal are adjusted for multiple test  
s.)
```

```
      kintamasis1 kintamasis2 kintamasis3 kintamasis4  
kintamasis1      0      0      0      0  
kintamasis2      0      0      0      0  
kintamasis3      0      0      0      0  
kintamasis4      0      0      0      0
```

To see confidence intervals of the correlations, print with the short=FALSE option

```
cr$r #tik koreliacijos koef
```

```
      kintamasis1 kintamasis2 kintamasis3 kintamasis4  
kintamasis1  1.0000000  0.7762114  0.9387647  0.9118741  
kintamasis2  0.7762114  1.0000000  0.7333548  0.7405589  
kintamasis3  0.9387647  0.7333548  1.0000000  0.9801467  
kintamasis4  0.9118741  0.7405589  0.9801467  1.0000000
```

```
cr$p #tik p reikšmės
```

```
      kintamasis1 kintamasis2 kintamasis3 kintamasis4  
kintamasis1 0.000000e+00 0.003294113 1.105696e-06 8.824795e-06  
kintamasis2 1.098038e-03 0.000000000 4.900595e-03 4.900595e-03  
kintamasis3 2.211392e-07 0.002838744 0.000000e+00 9.699642e-10  
kintamasis4 2.206199e-06 0.002450298 1.616607e-10 0.000000e+00
```

```
cr$ci.adj #pasikliautiniai intervalai
```

```
lower.adj upper.adj  
1 0.3040317 0.9422321  
2 0.7547765 0.9858240
```

```
3 0.6737128 0.9784378
4 0.3319304 0.9098924
5 0.2691133 0.9257082
6 0.9120663 0.9956378
```

```
#Regresija
#kintamasis3 - užimti gyventojai (mot.tūkst.)
#kintamasis4 - vidutinės ekvivalentinės neto pajamos
regr<-lm(kintamasis3~kintamasis4, data=macro, x=T)
summary(regr)
Call:
lm(formula = kintamasis3 ~ kintamasis4, data = macro, x = T)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-20.727  -8.310  -1.890   9.455  20.852
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.114e+03  2.468e+01  45.13 1.13e-15 ***
kintamasis4 1.923e-02  1.202e-03  16.00 6.21e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

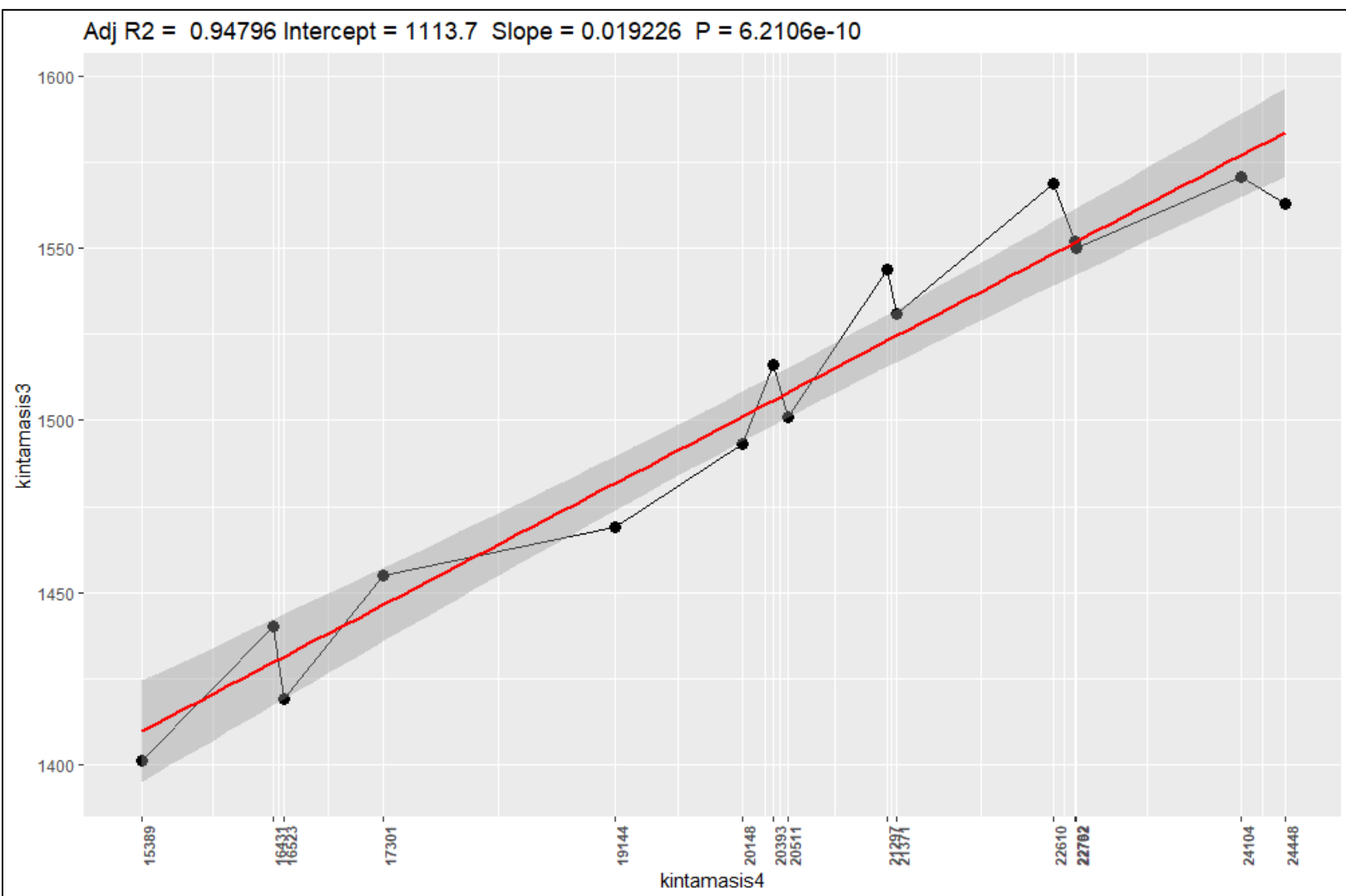
```
Residual standard error: 12.9 on 13 degrees of freedom
Multiple R-squared:  0.9517, Adjusted R-squared:  0.948
F-statistic: 256 on 1 and 13 DF, p-value: 6.211e-10
```

```
#regresijos grafikas
ggplotRegression <- function(fit,macro){

  require(ggplot2)

  ggplot(fit$model, aes_string(x = names(fit$model)[2], y = name
s(fit$model)[1])) +
    geom_line() +
    geom_point(size=3)+
    scale_x_continuous(breaks=fit$x[,2],labels=fit$x[,2])+
    stat_smooth(method = "lm", col = "red") +
    labs(title = paste("Adj R2 = ",signif(summary(fit)$adj
.r.squared, 5),
                                "Intercept =",signif(fit$coef[[1]],
5 ),
                                " Slope =",signif(fit$coef[[2]], 5)
,
                                " P =",signif(summary(fit)$coef[2,4
], 5)))+
    theme(axis.text.x = element_text(face="bold", size=8,
angle=90))
}

ggplotRegression(regr, macro)
```



(paaiškinimas: kintamasis3 - užimti gyventojai (mot.tūkst.;  
Kintamasis4 - vidutinės ekvivalentinės neto pajamos)

```
#####
#
#   prognozė: Užimti gyventojai (mot. tūkst.) nuo vid_neto_pajamu
2019-2021m.
#
#####
#suskaičiuoju "naujas" kintamojo4 reikšmes 2019-2021 m. naudodami
trendą
trnd1 <- lm(macro$kintamasis4 ~ Laikas, data=macro, x=T, y=T)
new.Laikas <- data.frame(Laikas = c(macro$Laikas, 2019:2021)) #prie
esamu metu pridėdi 3 naujus metus
new.kintamasis4 <- data.frame(kintamasis4=predict(trnd1, new.Laikas,
se.fit = TRUE)$fit)

#prognozė su pasikliautinaisiais intervalais
prognozel <- predict(regr, new.kintamasis4, se.fit = TRUE,
interval = "confidence")
```

```

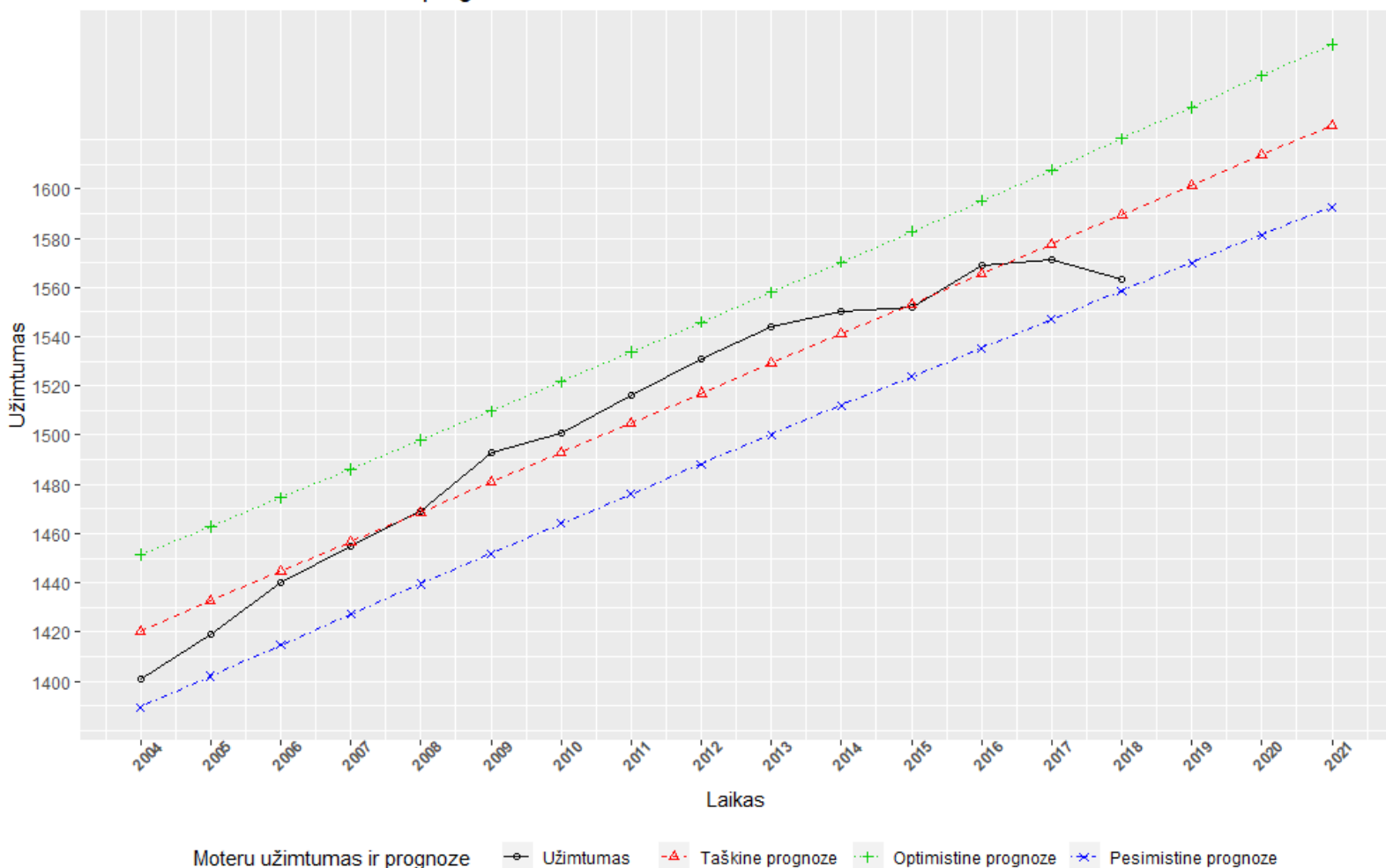
#Prognozė su prognozės intervalais
prognoze <- predict(regr, new.kintamasis4, se.fit = TRUE,
                    interval = "prediction")

#Duomenys grafikui (su prognozės intervalais)
g.macro <- data.frame(Laikas = new.Laikas,
                      kintamasis3 = c(macro$kintamasis3, rep(NA, 3)),
                      fitkintamasis3 = prognoze$fit[,1], #taškinė
                      opkintamasis3 = prognoze$fit[,3], #optimistinis
                      peskintamasis3 = prognoze$fit[,2]) #pesimistinis
mg.macro <- melt(g.macro, id="Laikas")

ggplot(mg.macro, aes(Laikas, value, color=variable, shape=variable,
                    linetype=variable)) +
  geom_line()+
  geom_point()+
  labs(x="Laikas", y="Užimtumas", title="Moterų užimtumas ir
užimtumo prognozė")+
  theme(legend.position="bottom")+
  scale_colour_manual(name = "Moterų užimtumas ir prognozė",
                      labels = c("Užimtumas", "Taškinė
prognozė", "Optimistinė prognozė",
                                "Pesimistinė prognozė"),
                      values = 1:4) +
  scale_shape_manual(name = "Moterų užimtumas ir prognozė",
                      labels = c("Užimtumas", "Taškinė prognozė",
                                "Optimistinė prognozė",
                                "Pesimistinė prognozė"),
                      values = 1:4) +
  scale_linetype_manual(name = "Moterų užimtumas ir prognozė",
                        labels = c("Užimtumas", "Taškinė
prognozė", "Optimistinė prognozė",
                                "Pesimistinė prognozė"),
                        values = 1:4)+
  scale_x_continuous(breaks=seq(2004, 2021, 1))+
  scale_y_continuous(breaks=seq(1400, 1600, 20))+
  theme(axis.text.x=element_text(face="bold", size="8",
angle=45))

```

## Moterų užimtumas ir užimtumo prognozė



**Komentaras:** Taigi iš grafiko matosi, jog moterų užimtumas, neturi vienos augimo ar mažėjimo tendencijos. Taip pat iš gauto grafiko galima matyti, jog 2009-2014 m. užimtumas buvo priartėjęs prie optimistinės prognozės. Tačiau po 2015 m. (žiūrint į užimtumo kreivę) moterų užimtumas sumažėjo (lyginant su 2009-2015m.) ir pradėjo sparčiai artėti prie pesimistinės prognozės.

```
#Duomenys grafikui (su pasikliautinaisiais intervalais)
gl.macro <- data.frame(Laikas = new.Laikas,
                        kintamasis3 = c(macro$kintamasis3, rep(NA, 3)),
                        fitkintamasis3 = prognozel$fit[,1], #taškinė
                        opkintamasis3 = prognozel$fit[,3], #optimistinis
                        peskintamasis3 = prognozel$fit[,2])

#pesimistinis

ggplot(gl.macro, aes(Laikas, fitkintamasis3)) +
  geom_line(color="blue") +
  geom_line(aes(Laikas, opkintamasis3), color=2) +
  geom_line(aes(Laikas, peskintamasis3), color=2) +
  geom_line(aes(Laikas, kintamasis3))+
  labs(x="Metai", y="Užimtumas", title="Moterų užimtumas ir
užimtumo prognozė")+
  scale_x_continuous(breaks=seq(2004, 2021, 1))+
```

```
scale_y_continuous(breaks=seq(1400, 1600, 20))+  
theme(axis.text.x=element_text(face="bold", size="8",  
angle=45))
```

**Moteru užimtumas ir užimtumo prognoze**

