

LINKÖPINGS UNIVERSITET

Visualisering 2

732G48 Introduktion till statistik och
dataanalys

Adrian Mansur, Thong Vinh Phat, Viet Tien Trinh, Duy Thai Pham

Höstterminen 2023

Innehåll

1. Inledning	1
2. Uppgifterna	2
2.1 Samband mellan ålder och lön vid NotReal AB	2
A)	2
B)	3
2.2 Antal bekräftade fall av COVID-19 i Sverige	4
A)	4
B)	5
2.3 Samband mellan olika husdjur	6
A)	6
3. Lärdomar, problem och övriga kommentarer	7
Bilaga	8
Kod uppgift 1B	8
Kod uppgift 2A	9
Kod uppgift 3	10

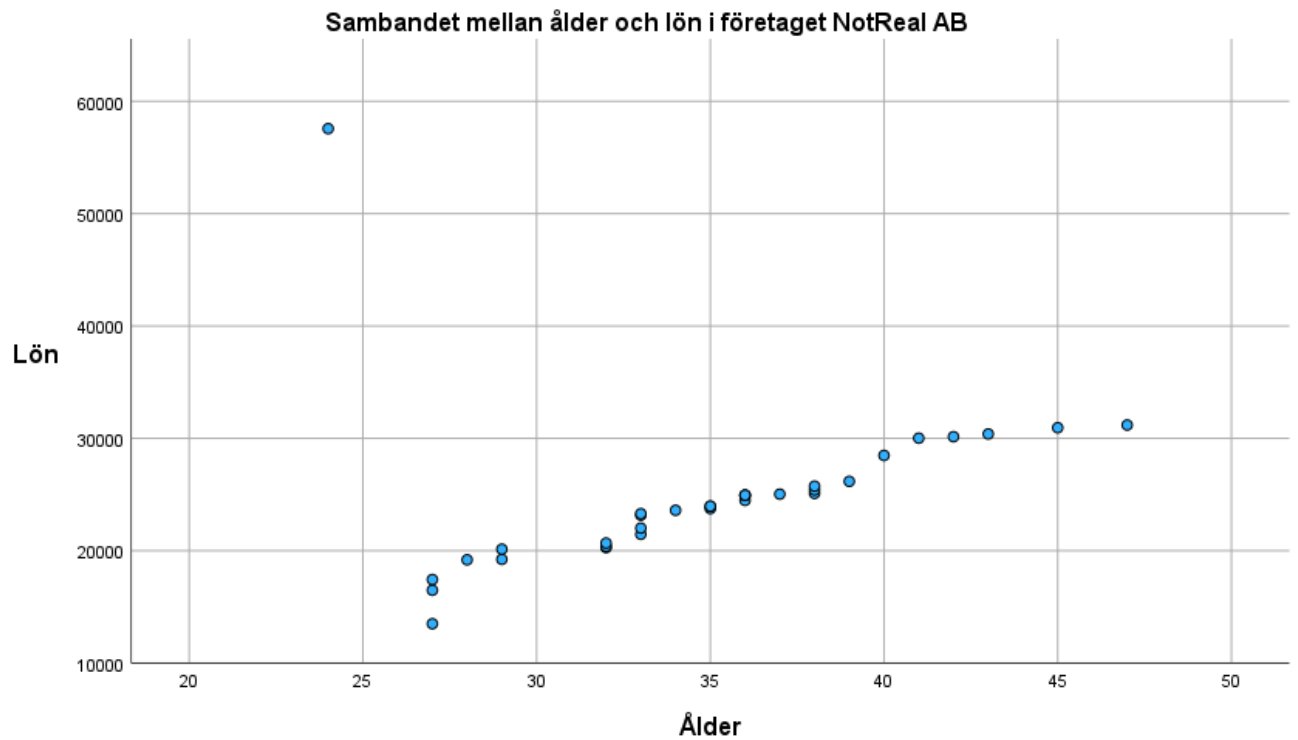
1. Inledning

Denna rapport handlar om gruppens arbete med tolkningen av olika diagram enligt visualisering 2 frågeställning.

2. Uppgifterna

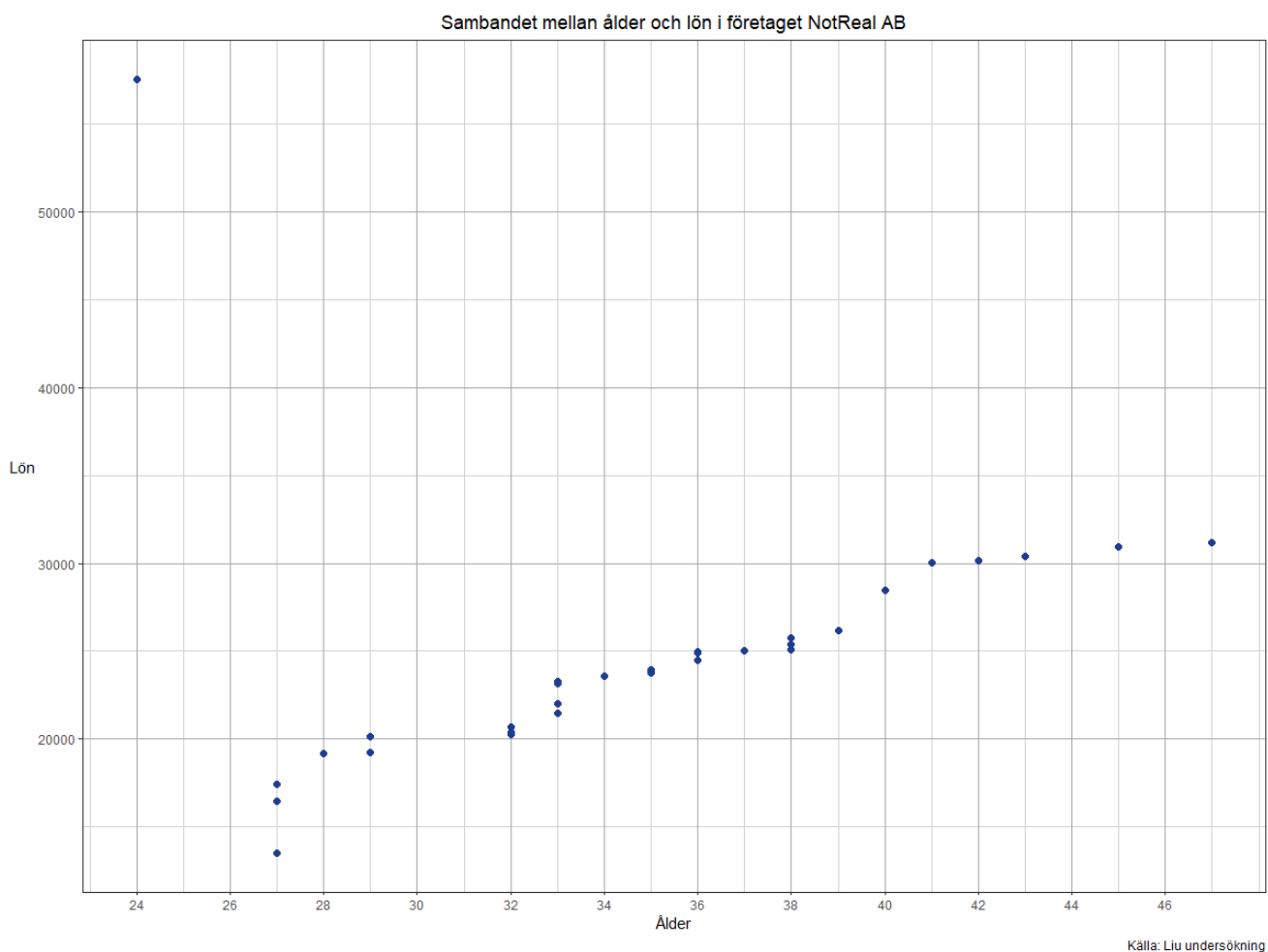
2.1 Samband mellan ålder och lön vid NotReal AB

A) Diagrammet nedan visar ett svagt positivt samband mellan ålder och lön. Lönen höjas i samband med ålder. Ju yngre anställda är, desto mindre lön de tjänar och vice versa.



Källa: Liu undersökning

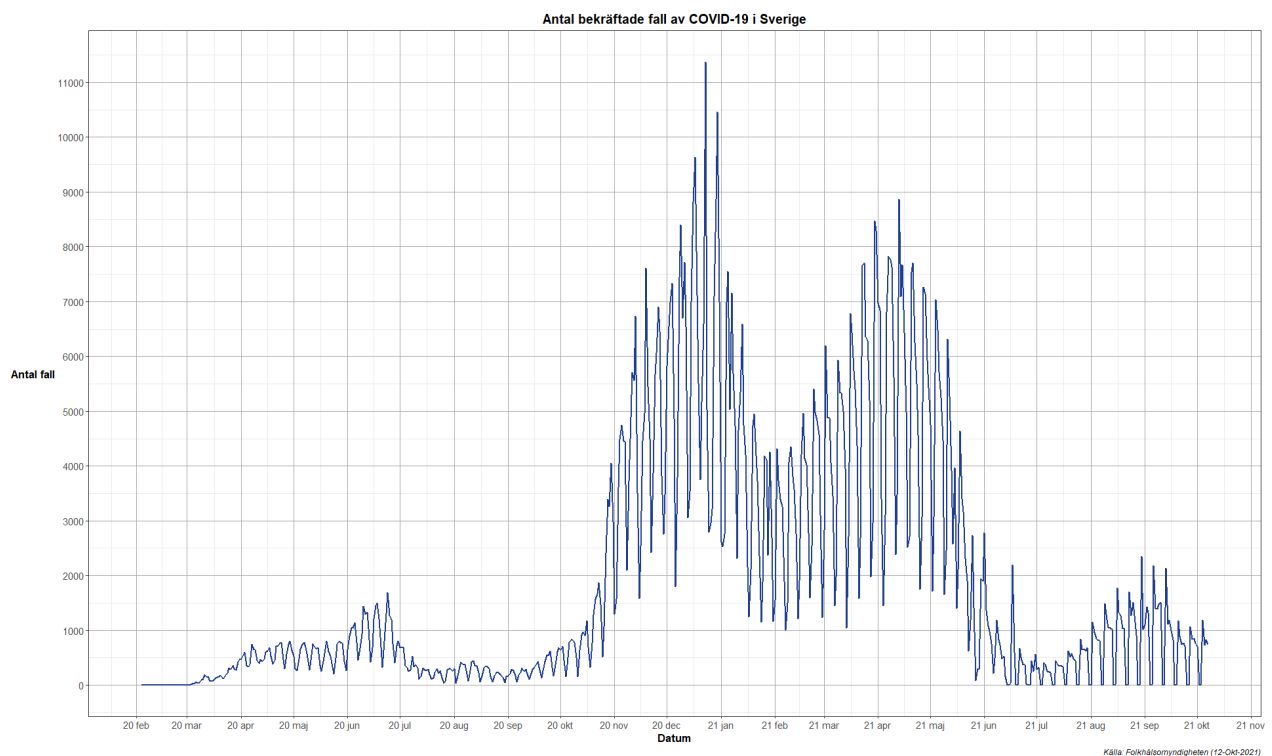
B) Diagrammet nedan visar ett svagt positivt samband. Dvs ju äldre anställda är, desto mer de tjänar. Eftersom löneökningen inte är särskilt stor därför blir det ett svagt och positivt samband. Det finns ett fall där en anställd i företag tjänar mer än alla andra men också är den yngsta i företaget. Denne anställd är ett extremt värde. Ett extremt värde har ingen koppling till diagrammets samband och behöver då en spekulatation. Spekulationen är att denna person anhåller den högsta positionen i företaget.



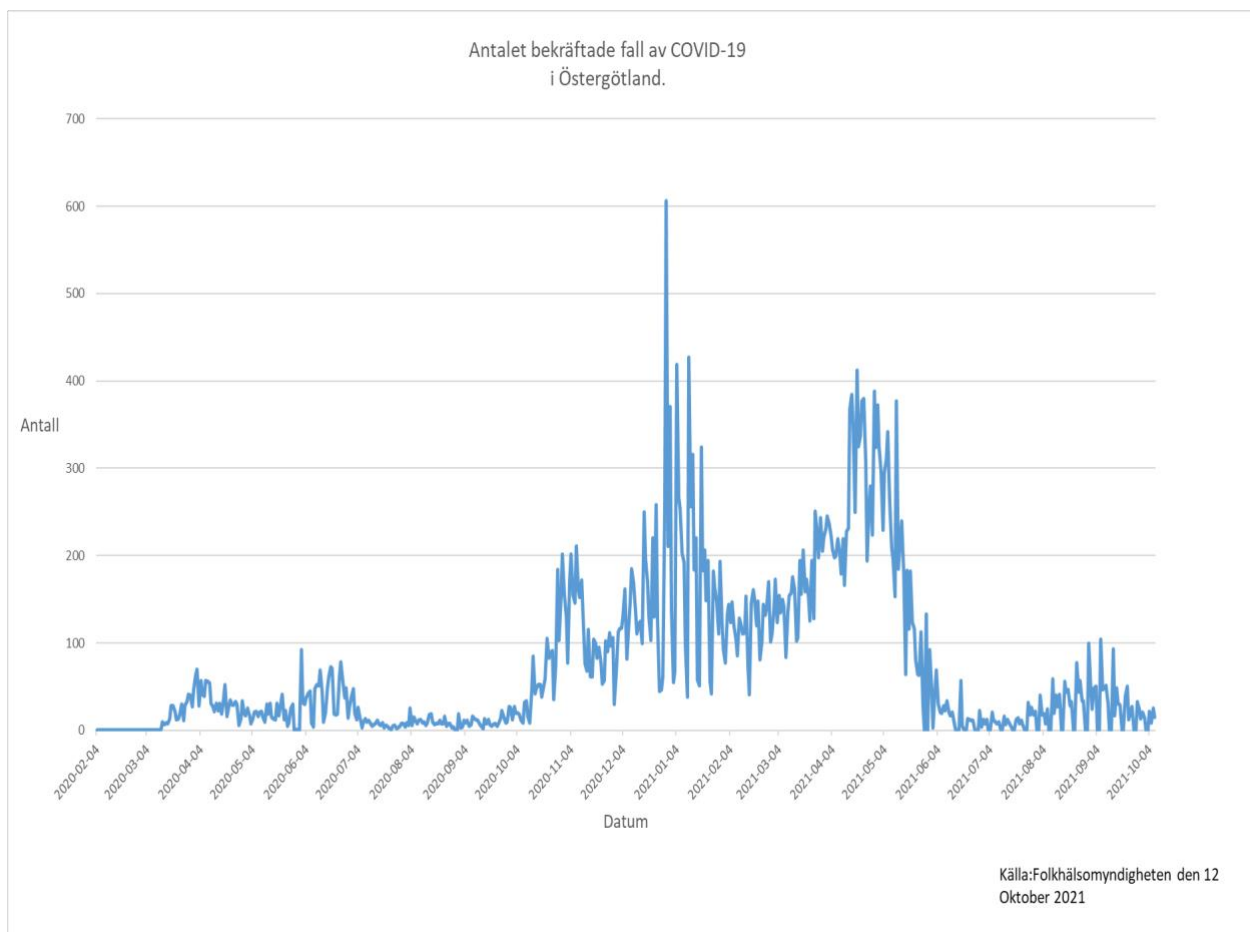
2.2 Antal bekräftade fall av COVID-19 i Sverige

A) Nedan beskriver ett linjediagram, nämligen en tidsserie antal bekräftade fall av COVID-19 i Sverige från ungefär början av februari 2020 till början av oktober 2021. Tidsserien har generellt ökat kraftigt i vissa perioder men i slutändan gått ner igen.

Det finns tydliga återkommande mönster som går att observeras i två vändpunkter, nämligen i mitten av december 2020 och i mitten av april 2021. Från mitten av oktober 2020 till mitten av december 2020 har serien kraftigt utvecklats från ungefär 1000 bekräftade fall av COVID-19 till nästan 11500 fall och sedan minskar kraftigt till ungefär 4000 fall i februari 2021. Det här mönstret av kraftig ökning och sedan minskning återkommer även i februari 2021, där serien har ökat från 4000 fall till nästan 9000 fall i mitten av april 2021 och efteråt minskat till ungefär 500 fall under juni och juli 2021. Förutom de två vändpunkterna som nämndes finns det också den tredje vändpunkten i diagrammet och den uppstår mellan de två topparna, alltså i februari 2021. Då vänder serien från en kraftig minskning till en ökning till en av topparna i mitten av april 2021.

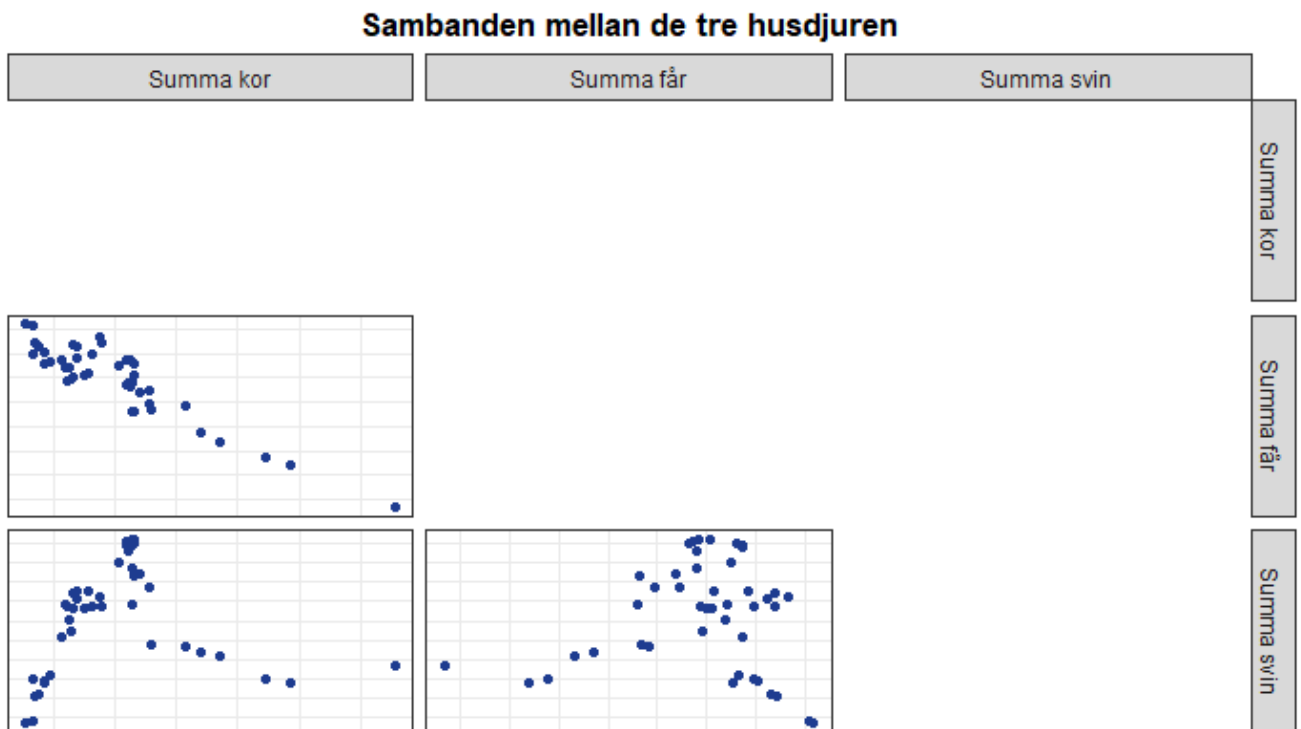


B) Nedanför beskrivs en linjediagram, en tidsserie av antal bekräftade fall av COVID-19 i Östergötland från början av februari 2020 till slutet av oktober 2021. Det har skett en kraftig ökning gradvis i en punkt mellan november 2020 och juni 2021 då det var mer än 100 bekräftade fall av COVID-19 i Östergötland, mellan denna punkt var januari 2021 toppad med 600 bekräftade fall av COVID-19 i Östergötland. Denna tidsserie har inte så många vändningar efter toppningen då det har blivit mindre bekräftade fall och sedan ökning med bekräftade fall i början av april 2021 till slutet av maj 2021.



2.3 Samband mellan olika husdjur

A) I detta R Studio punktdiagrammatris visas parvisa samband mellan de tre olika husdjuren. I första ruta av den andra raden är det samband mellan kor och får, man kan se detta samband är negativt eftersom de flesta punkter ligger nära varandra och lutningen är neråt. Nedan i tredje raden är två sambandet mellan svin – kor och svin - får, man kan se de punkterna ligger i spritt i diagram och de gör inte en tydlig koppling och vi kan bedöma de är ganska svagt men de inte skapa ett negativt samband.



Källa: SCB Undersökning

3. Lärdomar, problem och övriga kommentarer

I denna laboration uppgift skapade vår grupp olika diagram via några hjälpmedel som SPSS, Excel och R studio, samtidigt hänvisa vi till vissa kunskaper från föreläsning för att hantera frågor och använde datamaterial som publicerats i Lisam. Vi upplever inga svårigheter under detta moment.

Bilaga

Kod uppgift 1B

#Steg 0: Aktivera ggplot2, importera datamaterialet från excel.

```
library(ggplot2)
```

#Steg 1: Skriv datamaterialets namn i ggplot och döper x, y till de variabler från excel som vi vill undersöka.

```
ggplot(VIS1_Lön) +
```

```
  aes(x = Ålder, y = Lön) +
```

#Steg 2: Skapa ett spridningsdiagram med funktionen geom_point samt förstora punkterna med size och omfärgad punkterna med color.

```
  geom_point(size = 2, color = "#1E3C90") +
```

#Steg 3: Lägga till theme_bw() för att snygga till diagrammet. Denna funktion ändrade bakgrunden till vit från en grå bakgrund. Den läggs också till en border runt diagrammet.

```
  theme_bw() +
```

#Steg 4: Denna steg används funktion theme() för att justera positionen på diagrammets titel respektive y-axeln position. Ytterligare grid lines på både y och x axlarna var förtydligare genom att mörka till färgerna.

```
  theme(axis.title.y =
```

```
    element_text(angle = 0,
```

```
      hjust = 1,
```

```
      vjust = 0.5),
```

```
  plot.title = element_text(hjust = 0.5),
```

```
  plot.subtitle = element_text(hjust = 0.5),
```

```
  panel.grid.major.x = element_line(color = "darkgray"),
```

```
  panel.grid.minor.x = element_line(color = "lightgray"),
```

```
  panel.grid.major.y = element_line(color = "darkgray"),
```

```
  panel.grid.minor.y = element_line(color = "lightgray")) +
```

#Steg 5: Det är viktigt att ge titlarna på diagrammet, dvs x, y axlar respektive huvudtitel det namnet de ska heta. Källa ska också anges och allt detta är gjort med labs funktionen.

```
  labs(title = "Sambandet mellan ålder och lön i företaget NotReal AB",
```

```
x = "Ålder",
```

```
y = "Lön",
```

```
caption = "Källa: Liu undersökning") +
```

#Steg 6: För att förtydliga x-axeln's värde, har funktionen `scale_x_continuous` använts. De anställda i datamaterialet är från 24–47 år gammal, på diagrammet visas 24 som den första värden på x-axel, sluta med 46 och det brytes varje två värden.

```
scale_x_continuous(breaks = seq(24,47,2))
```

Kod uppgift 2A

```
#Vi använder packet ggplot2.
```

```
library(ggplot2)
```

```
#Vi använder data som importera från excel kalla vi det för "data_dödsfall" för att rita en  
#linjediagram med två variabel "Statistikdatum" och "Totalt antal fall".
```

```
ggplot(data_dödsfall) +  
  geom_line(aes(x = Statistikdatum,  
                y = Totalt_antal_fall,  
                group = 1),  
            size = 1,  
            color = "#1E3C90") +
```

```
#Bakgrund blir vita och border läggs till.
```

```
theme_bw() +
```

```
#Skala om x axeln, avstånd mellan två datum är en månad.
```

```
scale_x_datetime(date_breaks = "1 month", date_labels = "%y %b") +
```

```
#Vi kalla variabel i y axeln är "Antal fall", i x axeln är "Datum", vi lägga till caption "Källa:
```

```
#Folkhälsomyndigheten" och title "Antal bekräftade fall av COVID-19 i Sverige".
```

```
labs(y = "Antal fall",  
     x = "Datum",  
     caption = "Källa: Folkhälsomyndigheten (12-Okt-2021)",  
     title = "Antal bekräftade fall av COVID-19 i Sverige") +
```

```

#Redigera text: stolek, position, vinkel, djärv.
theme(plot.title = element_text(hjust = 0.5,
                                size = 15,
                                face = "bold"),
      axis.title.y = element_text(face = "bold",
                                   angle = 0,
                                   vjust = 0.5,
                                   size = 13),
      axis.title.x = element_text(face = "bold",
                                   size = 13),
      plot.caption = element_text(size = 9,
                                   face = "italic"),
      axis.text.x = element_text(size = 11),
      axis.text.y = element_text(size = 11),

#färgad major line av x och y är " darkgray ".
      panel.grid.major.x = element_line(color = "darkgray"),
      panel.grid.major.y = element_line(color = "darkgray")) +

#skala om y axeln med längden från 0 till 12000 och avstånd mellan två värde
#är 1000.
      scale_y_continuous(breaks = seq(from = 0, to = 12000, by = 1000))

```

Kod uppgift 3

#Vi använder paket "GGally"

```
library(GGally)
```

#Vi använder data som importera från excel "VIS2_Husdjur_Sep2023" för att rita en
#punktdiagramsmatris. Rensa bort rutor ovan, diagonal, axelmärke och lägga till title
#"Sambanden" mellan de tre husdjuren.

```

ggpairs(data = VIS2_Husdjur_Sep2023,
        upper = "blank",
        diag = "blank",
        axisLabels = "none",
        title = "Sambanden mellan de tre husdjuren",

```

#Tilldela färg .

```
mapping = aes(color = "färg")) +
```

```
scale_color_manual(values = "pink") +  
#lägg till caption "Källa: SCB Undersökning"  
labs(caption = "Källa: SCB Undersökning") +  
#Bakgrund blir vita. Grafens border läggs till.  
theme_bw() +  
#Redigera text: stolk, position, djärv eller kursiv.  
theme(plot.title = element_text(size = 13,  
                                hjust = 0.5,  
                                face = "bold"),  
       plot.caption = element_text(face = "italic",  
                                   size = 11))
```