

LINKÖPINGS UNIVERSITET

Visualiseringsrapport grupp 11

732G48 Introduktion till statistik och
dataanalys

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

Höstterminen 2023

Innehåll

1. Inledning	1
2. Uppgifterna	2
2.1 Partisimpatier	2
A)	2
B)	3
C)	4
2.2 Lönefördelning på företaget NotReal AB.....	5
A)	5
B)	7
3. Lärdomar, problem och övriga kommentarer	8
Bilaga	9
RStudios kod, uppgift 1B	9
RStudios kod, uppgift 2B	10

1. Inledning

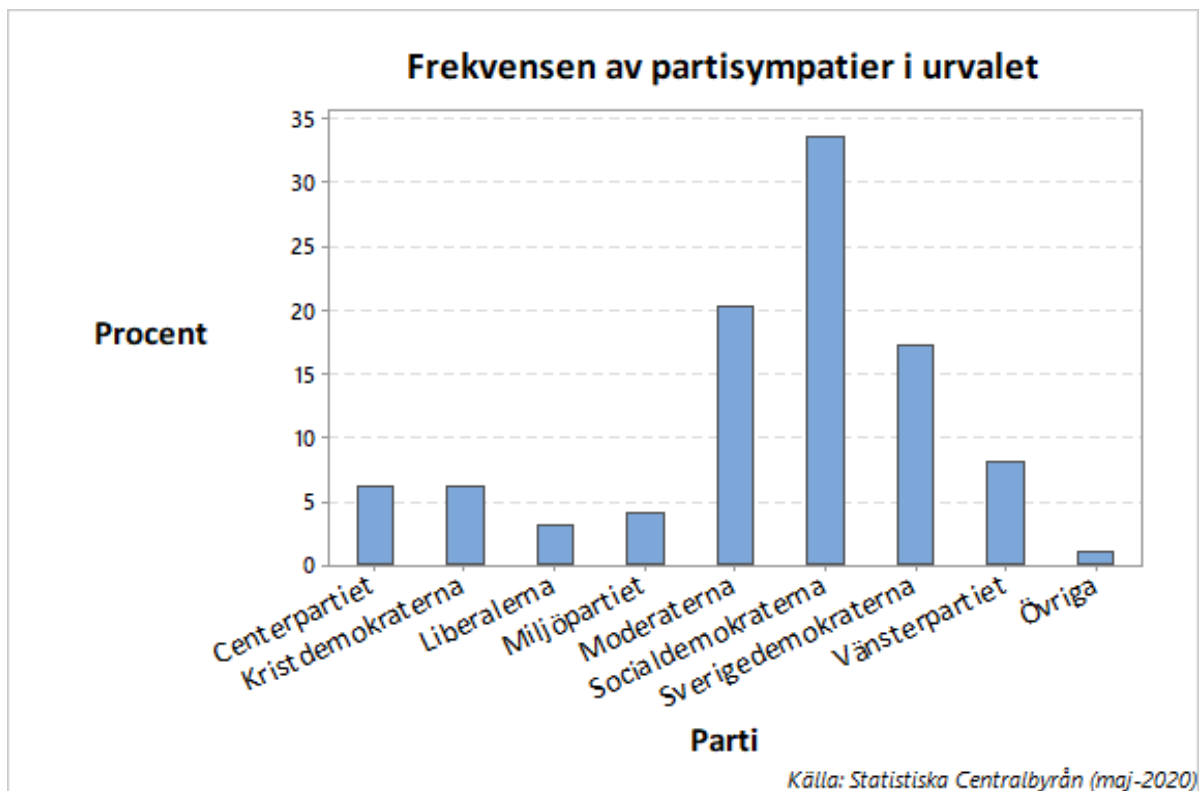
I denna rapport ska gruppen förklara olika diagram som skapades i programvaror R Studio, SPSS och Minitab. Dessa diagram är möjliggjorts med hjälp av datamaterialet som hämtades från Statistiska Centralbyrån och en påhittade databasen på företaget "NotReal AB".

2. Uppgifterna

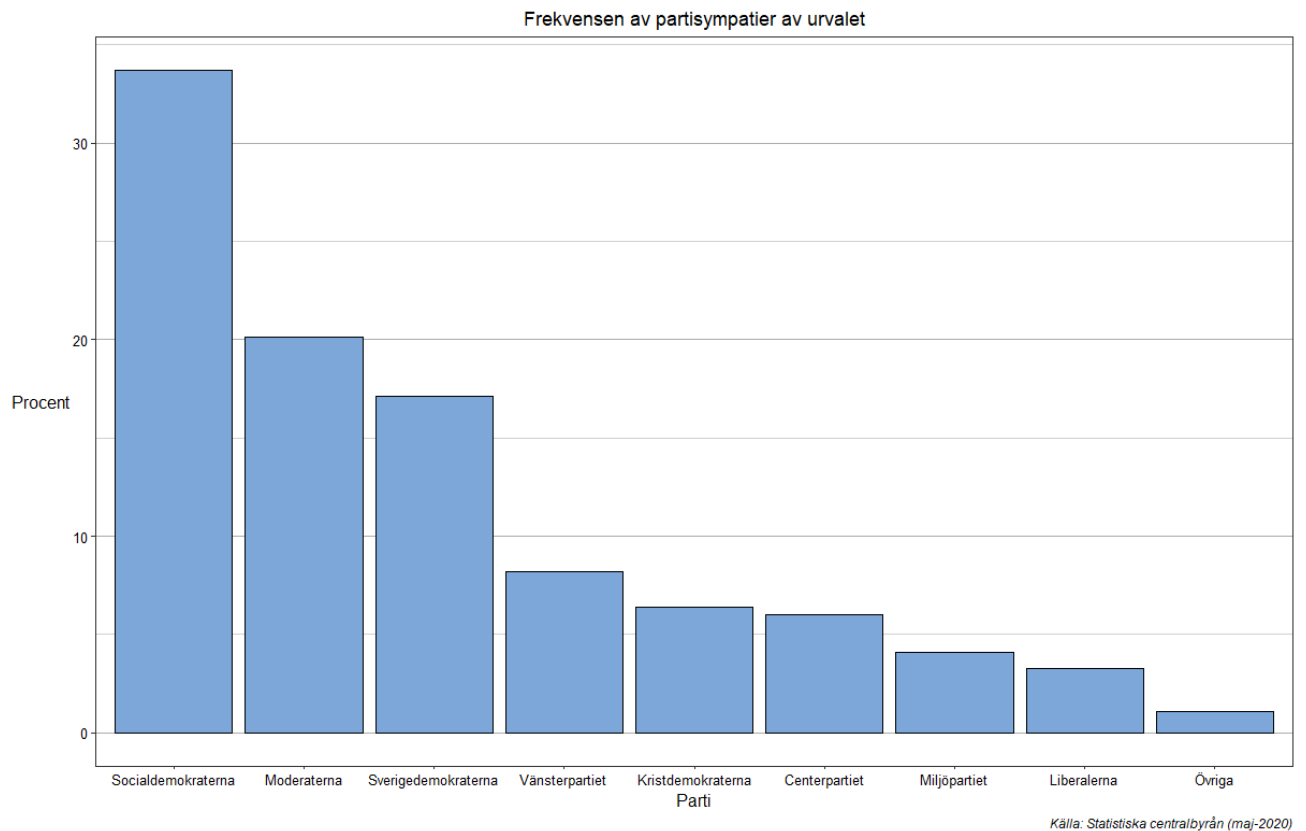
2.1 Partisympatier

A) Diagrammet nedan visar en fördelning av olika partisympatier under urvalet 2020.

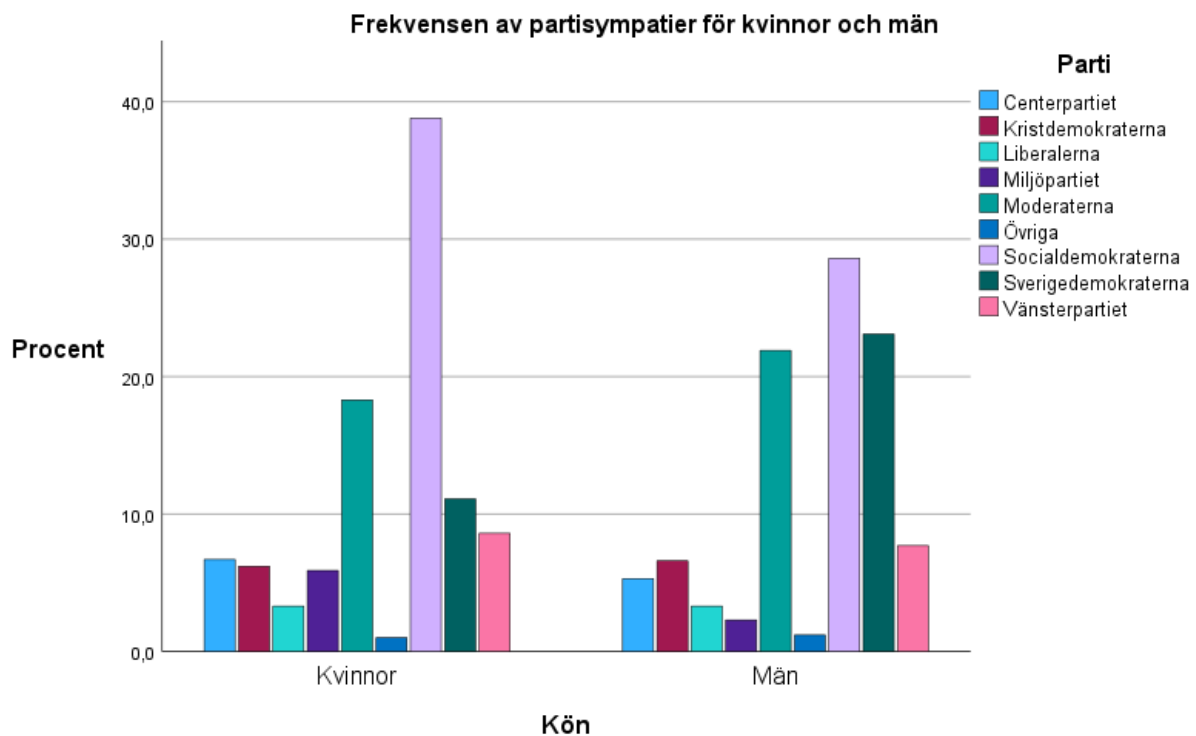
Socialdemokraterna har mer än 30% röster från partisympatisörer. Moderaterna har drygt över 20% röster. Sverigedemokraterna har ungefär 17% röster. Vänsterpartiet ligger runt 8% med röster. Centerpartiet och Kristdemokraterna har ett ungefärligt jämt procentig på 6%. Liberalerna har ungefär 3% röster. Miljöpartiet har 4% röster och övriga partirepresentanter äger ungefär 1% av röster från partisympatier.



B) Stapeldiagrammen nedan visar olika partier och deras totala procentuella röster under urvalet 2020 i fallande ordning. Socialdemokraterna äger strax under 35% röster, men är ledande parti jämför med andra partier såsom Moderaterna som har lite drygt över 20%, eller Sverigedemokraterna som äger ungefär 17% av de totala rösterna. De resterande partier har alla under 10% röster. De är bland annat Vänsterpartiet med cirka 8%, Kristdemokraterna och Centerpartiet med cirka 6%, Miljöpartiet med cirka 4%, Liberalerna med cirka 3% och övriga partier äger ungefär 1% av rösterna.



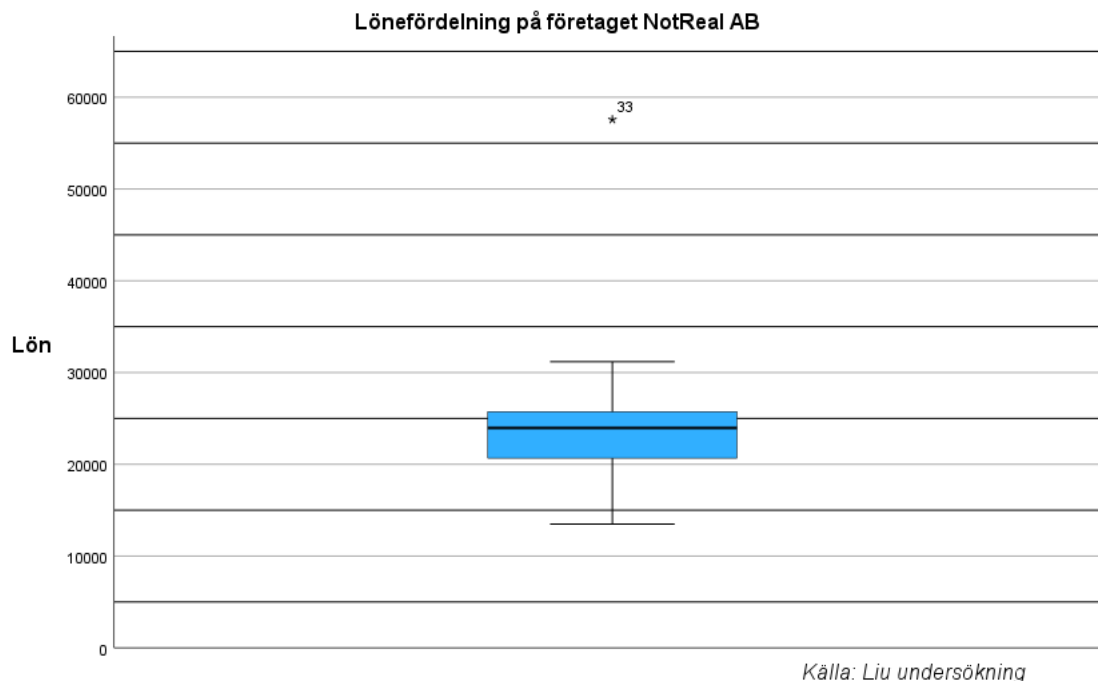
C) Den grupperat stapeldiagram nedan visar frekvensen av partisympatier för kvinnor och män. Socialdemokraterna får mest röster från bägge kön, ungefär 38% av röster från kvinnor och 28% rösterna från män. Moderaterna får ungefär 18% av rösterna från kvinnor och cirka 22% från män. Sverigedemokraterna får ungefär 23% av sina röster från män och cirka 11% från kvinnor. Vänsterpartiet får ungefär 8% av sina röster från kvinnor och cirka 7% från män. Centerpartiet får ungefär 6% av sina röster från kvinnor och ungefär 5% från män. Kristdemokraterna får ungefär 6% av röster från män och cirka 5% från kvinnor. Miljöpartiet får mest av sina röster från kvinnor som motsvarar ungefär 5% och ungefär 2% av rösterna från män. Liberalerna får ungefär lika jämna röster från både män och kvinnor, cirka 3% av rösterna från båda kön. Övriga partier har likvärdiga röster på ungefär 1% från män respektive kvinnor.



Källa: Statistiska centralbyrån (maj-2020)

2.2 Lönefördelning på företaget NotReal AB

A) diagrammet nedan redovisas en spridning av anställdas olika löner på företaget NotReal AB. Data med olika värden i storleksordning importeras från Excel till SPSS och där har den 33:e observationen det största värdet, vilket innebär högsta lönen bland andra anställda i NotReal AB. När lådagrammet skapas finns det en punkt med nummer 33 som ligger utanför. Punkten är därmed ett extremvärde. Lådagrammet visar även var medianen och kvartilerna Q1 och Q3 ligger ungefär. Utifrån lådagrammet kan vi se att medianen är lite mindre än 25000kr och större än 22500kr. Den första kvartilen, Q1 är 21000kr. Den tredje kvartilen, Q3 är ungefär 26000kr. Extrempunkten ligger på cirka 57500kr. Dessutom har vi en MIN-punkt på ungefär 13000kr och en MAX-punkt på ungefär 31000kr. MIN-punkten visar den observationen som har lägst värde, vilket innebär att den lägsta lönen som en anställd på NotReal AB tjänar är 13000kr och MAX-punkten visar tvärtom att den näst högsta lönen som en annan anställd tjänar är 31000kr. Det finns även ett extremvärde, vilket innebär att en anställd tjänar högst 57500kr och denna observation ligger därmed långt utanför lådagrammets räckvidd.



Med hjälp av informationen ovanpå kan vi även räkna ut vad variationsvidd och kvartilavstånd ungefär blir. Variationsvidd som är ett mått på spridningen för hela datamaterialet kan beräknas genom att räkna fram skillnaden mellan den största och minsta observationen. I vårt fall blir det skillnaden mellan MAX-punkt och MIN-punkt. Kvartilavstånd som är ett mått på spridning på de mittersta 50% av observationen kan beräknas genom att räkna fram skillnaden mellan Q1 och Q3.

Variationsvidd =

$$31000kr - 13000kr = 18000kr$$

Här kan vi alltså tolka att skillnaden mellan den anställde som tjänar lägst och den anställde som tjänar näst högst är 18000kr.

Kvartilavstånd, IQR =

$$26000kr - 21000kr = 5000kr$$

Detta innebär att skillnaden mellan den anställde som tjänar lägst och den anställde som tjänar högst inom de mittersta 50% av observationen är 5000kr.

Med hjälp av beräkningar nedan kan det exakta värdet av medianen, kvartilerna Q1 och Q3 fås fram.

Positionen av medianen = $0.5 * (33+1) = 17$

Detta medför att medianen är observationen nummer 17, vilket är 23960kr.

Positionen av den första kvartilen, $Q1 = 0.25 * (33+1) = 8.5$

Det innebär att vi måste beräkna ett medelvärde mellan värden av observationen nummer 8 och 9.

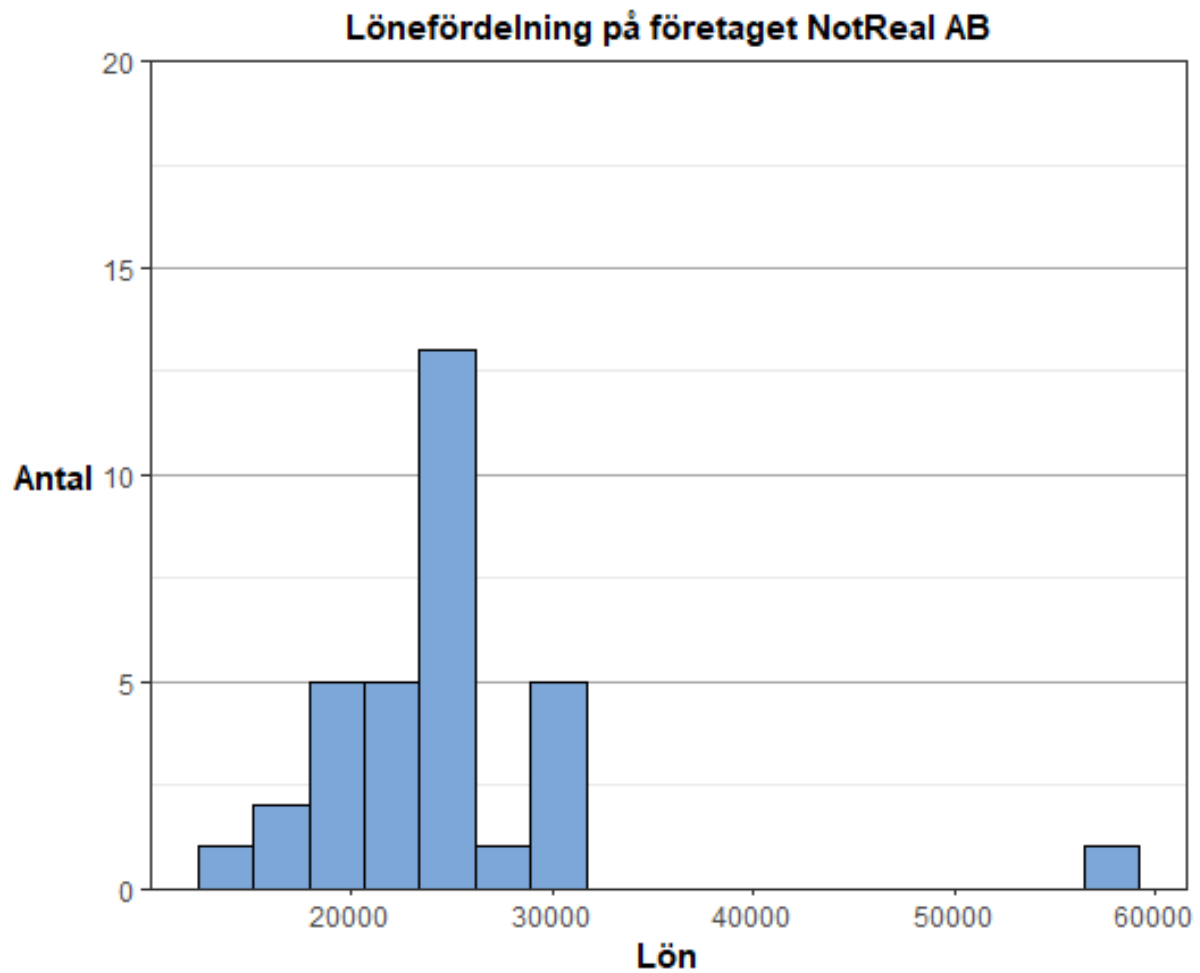
$$Q1 = \frac{20383\text{kr} + 20673\text{kr}}{2} = 20528\text{kr}$$

Positionen av den tredje kvartilen = $0.75 * (33 + 1) = 25.5$

Detta innebär att vi måste beräkna ett medelvärde mellan observationen nummer 25 och 26.

$$Q3 = \frac{25740\text{kr} + 26168\text{kr}}{2} = 25954\text{kr}$$

B) Diagrammet nedan visar lönefördelningen för anställda i företaget NotReal AB. Detta är inte ett symmetriskt diagram och har bara en topp. De flesta anställda som får löner sträcker sig från 18000kr till 32000kr. 3 anställda som får sin lön under 18000kr och bara 1 anställd får nästan 60000kr i lön, det är också extremvärdet av diagrammet.



Källa: Liu undersökning

3. Lärdomar, problem och övriga kommentarer

Genom denna uppgift lärde vi oss hur man skapa ett diagram via Minitab, SPSS eller R studio, samt hur man arbeta i grupp i Lab. Vi delar upp uppgifter i många delar och varje person tar på sig en del och syntetiserar slutligen all information om uppgifter tillsammans. Generellt sett hade vi inte så många svårigheter när vi utför denna uppgift, det var bara lite svårt att kommunicera med varandra när vi inte var i labbet.

Bilaga

RStudios kod, uppgift 1B

```
#Vi använde packet ggplot2
library(ggplot2)
#Vi tala om vilken data som används och skapa ett stapeldiagram med inuti staplar
#är "#7da7d9" och gräns omkring av staplarna är svart.
#Samtidigt sortera vi fallande ordning med reorder()funktion.
ggplot(data = VIS1_Partisympatier, aes(x = reorder(Parti, -Procent), y = Procent)) +
  geom_bar(stat = "identity",
    fill = "#7da7d9",
    color = "black") +
  #bakgrund är vita.
  theme_bw() +
  #Vi tar bort major_line och minor_line av x och färgad major av y med color ="grey68" och
  #Minor_line av y med color = "grey82".
  theme(panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank(),
    panel.grid.major.y = element_line(color = "grey68"),
    panel.grid.minor.y = element_line(color = "grey82"),
  #Vi ändra textformat av caption till "italic" och skala om titles position.
    plot.caption = element_text(face = "italic"),
    plot.title = element_text(hjust = 0.5),
  #Vi skala om position, storlek, vinkel och färg av texten i x och y axel
    axis.title.y = element_text(angle = 0, vjust = 0.5, size = 12),
    axis.title.x = element_text(size = 12),
    axis.text = element_text(color = "black", size = 10)) +
  #Vi nämn x är Parti, caption är Källa: Statistiska centralbyrån (maj-2020) och
  # title är Frekvensen av partisympatier av urvalet
  labs(x = "Parti",
caption = "Källa: Statistiska centralbyrån (maj-2020)",
    title = "Frekvensen av partisympatier av urvalet")
```

```
#Vi använder packet "ggplot2"
library(ggplot2)

#Vi tala om vilken data som ska användas
P = ggplot(VIS1_Lön) + aes(Lön)
#Vi skapa ett histogram med en bredd som motsvarar 17 staplar och inuti varje staplarna
#är "#7da7d9" och gräns omkring våra staplar är svarta.
P = P + geom_histogram(fill="#7da7d9",
                        color="black",
                        bins = 17) +
#background är vita.
theme_bw() +
#Vi skalar om och tar bort den tomma ytan under staplarna och
#expanderar y-axeln för att få största värde är 20 i diagramet.
scale_y_continuous( expand = c( 0 , 0 ),
                    limits = c ( 0 , 20 ))+
#Vi nämn Y-axeln är "Antal" och skapa en caption är "Källa: Liu
#undersökning och en title som är "Lönefördelning på företaget
#NotReal AB".
labs(y="Antal",
     caption = "Källa: Liu undersökning",
     title = "Lönefördelning på företaget NotReal AB") +

#Vi skalar om position, storlek, djärvare bokstav och vinkel av
#title och texten i x och y axel.
theme(plot.title = element_text(hjust = 0.5,
                                face="bold",
                                size = 11),
      axis.title.x = element_text(face ="bold"),
      axis.title.y = element_text(face ="bold",
                                  angle = 0,
                                  vjust = 0.5),
      #Vi tar bort major_line och minor_line av x och färgad major_line av Y är
      #darkgray.
      panel.grid.major.x = element_blank(),
      panel.grid.minor.x = element_blank(),
      panel.grid.major.y = element_line(color = "darkgray"))
#print P.
P
```