

Projekt 732G33

Författare: Vinh Phat Thong och Viet Tien Trinh.

LiU-id: vint530 och viet933

Grupp: 13

Dataanalys av kommundata

Vi arbetar med en dataset som innehåller information om befolkning och utbildningsnivå för olika regioner i Sverige. De variabler som ingår i vår analys är:

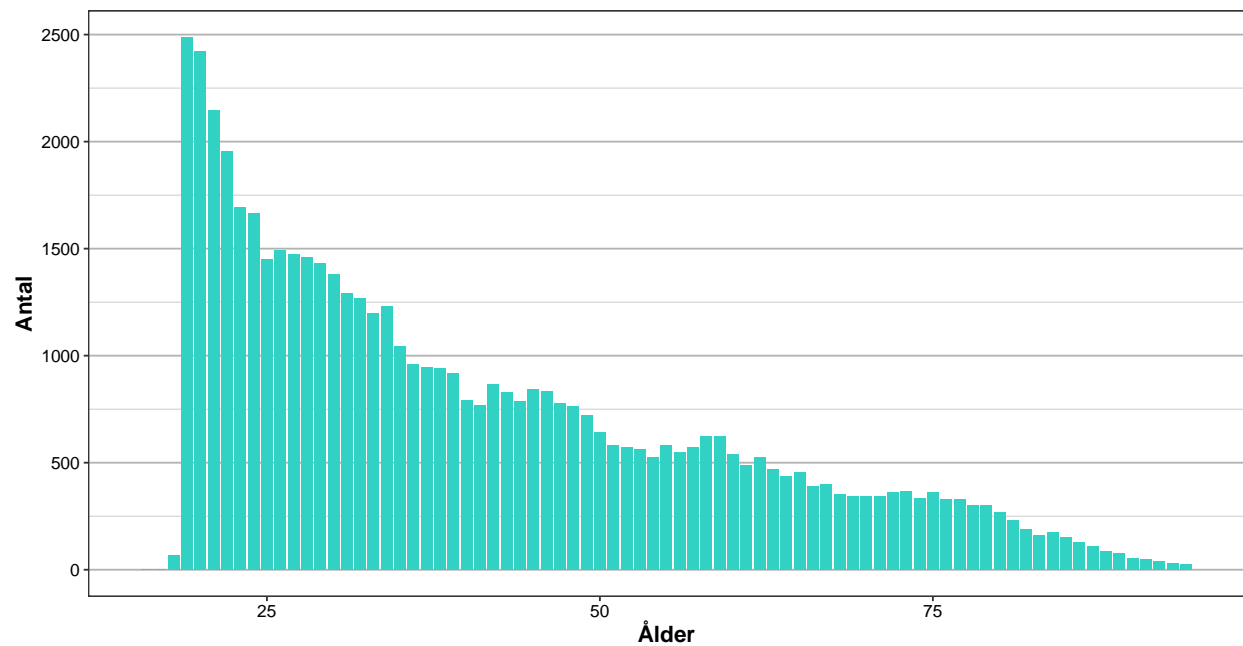
- Region: Namnet på regionen eller kommunen.
- Befolkning_16_to_94: Totalt antal invånare i åldrarna 16 till 94 år i respektive region.
- Antal_per_10000: Antal invånare per 10,000 invånare i respektive region.
- Utbildningsnivå: Den högsta utbildningsnivån för majoriteten av invånarna i regionen, i detta fall "3-årig gymnasieskolan".
- År: Årtalet för datainsamlingen, i detta fall är alla observationer från år 2023.
- Total antal invånare: Totalt antal invånare i respektive region.

Table 1: Tabell 1: 5 kommuner och variabler

region	Befolkning_16_to_94	antal_per_10000	utbildningsnivå	år	total_antal_invånare
Linköping	27906	1666.985	3-årig gymnasieskolan	2023	167404
Lund	16189	1242.555	3-årig gymnasieskolan	2023	130288
Malmö	54322	1500.057	3-årig gymnasieskolan	2023	362133
Norrköping	27420	1888.911	3-årig gymnasieskolan	2023	145163
Stockholm	142552	1441.458	3-årig gymnasieskolan	2023	988943

Uppgift 1

Fördelning av antal per 10000 av personer som slutför en 3-årig gymnasieskolan efter ålder i Malmö kommun



Källa: SCB(2023)

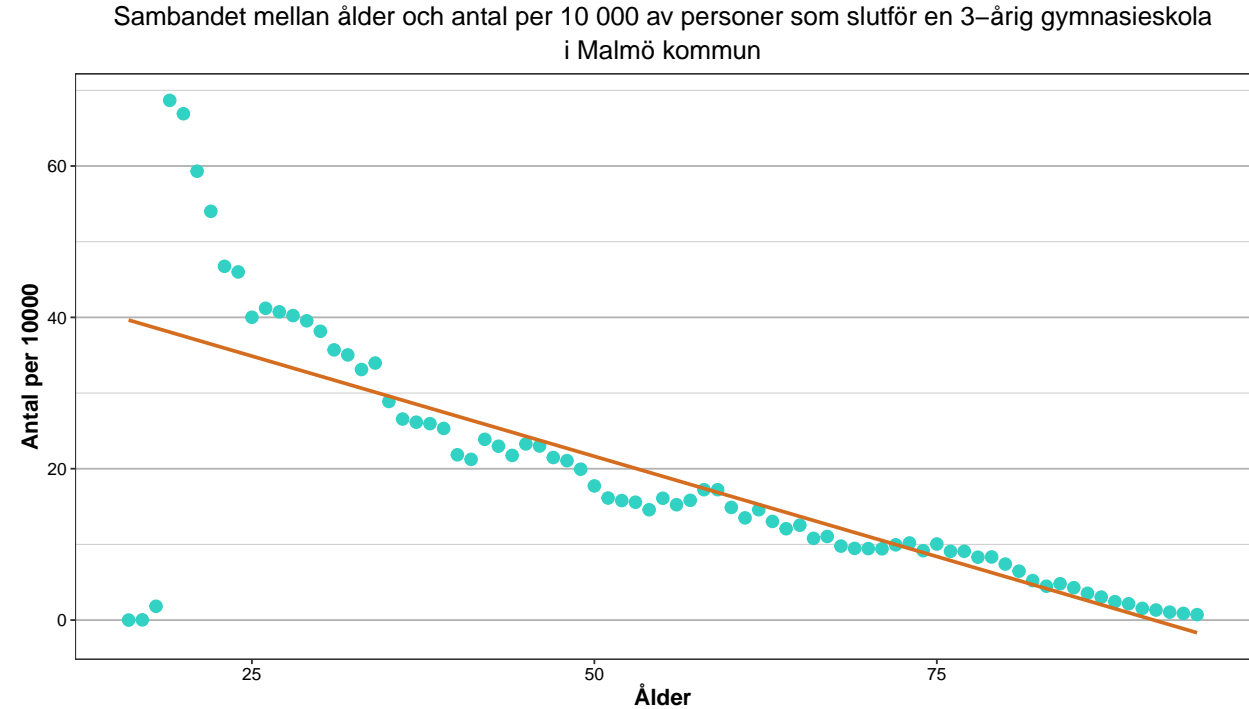
Ovanför beskrivs en stapeldiagram av fördelning av antal personer som slutför 3-årig gymnasieskolan efter ålder på kommunalsnivå med x-axel är ålder och y-axel är antal. Det har skett ett karftig minskning från vänster till höger, dvs att andel personer som slutför 3 årig +ymnasieskolan största finns bland ungdomar under 20 år gammal och minskning gradvis till 94 år gammal.

Uppgift 2



Diagrammet ovan visar fördelning av antal per 10000 av personer som slutför 3-årig gymnasieskolan på histogram med x-axel är antal per 10000 och y-axel är frekven. Detta är inte ett symmetriskt diagram och har bara en topp. De flesta andel av personer som slutför 3-årig gymnasieskolan sträcker sig från 0 per 10000 till 20 per 10000. Man kan lätt hitta 3 streckade linjer, de motsvara respektive första kvartiler, median och tredje kvariler.

Uppgift 3



Diagrammet ovan visar ett starkt negativt samband mellan ålder och antal per 10000 med x-axel är ålder och y-axel är andelen. Regressionlinjer visar en nedåtgående trend från vänster till höger, ge, man kan säga ju öka ålder, desto minskar andelen av personer som slutför 3-årig gymnasieskolan.

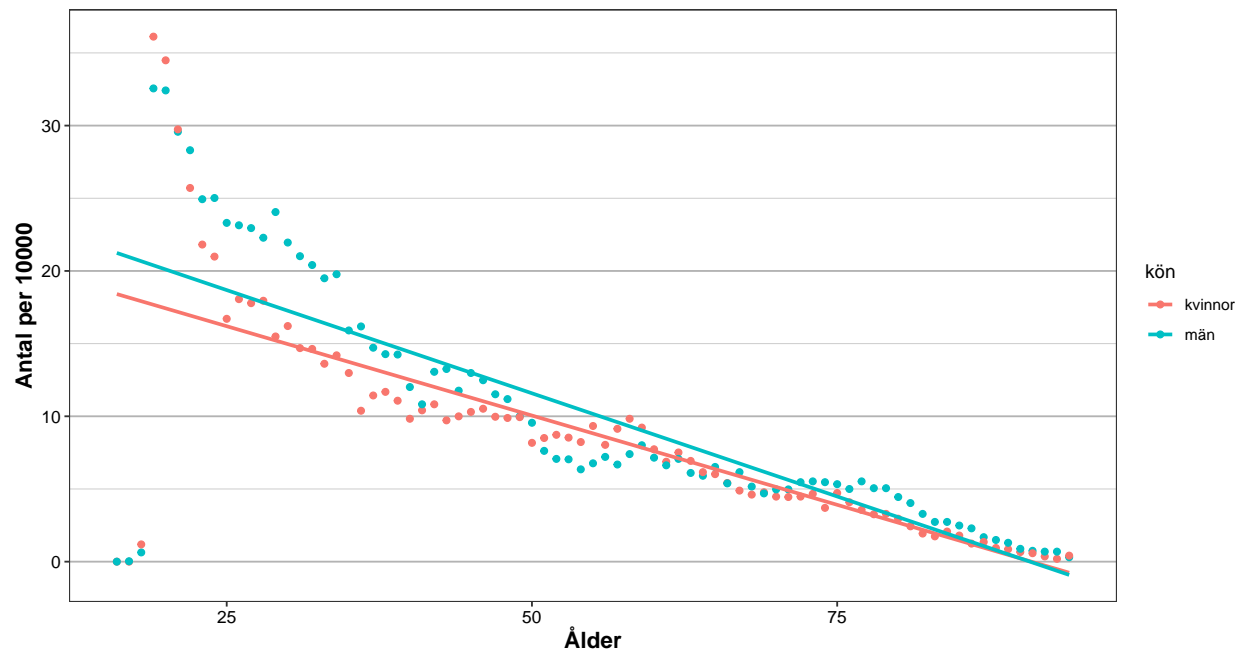
Uppgift 4

```
##
## Pearson's product-moment correlation
##
## data: x and y
## t = -14.555, df = 156, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.8180984 -0.6837826
## sample estimates:
## cor
## -0.7588984
```

Med 95% säkerhet kan man säga att korrelationen mellan ålder och andel per 10000 invånare, som har klarat 3 års gymnasial utbildning i Linköpings kommun år 2023 ligger någonstans mellan -0.787 och -0.547. Äldre åldersgrupper tenderar att ha lägre andelar som klarat 3 års gymnasial utbildning jämfört med yngre åldersgrupper.

Uppgift 6 (a)

Sambandet mellan ålder och antal per 10000 av personer som slutför 3-årig gymnasieskolan
i Malmö kommun



Källa: SCB(2023)

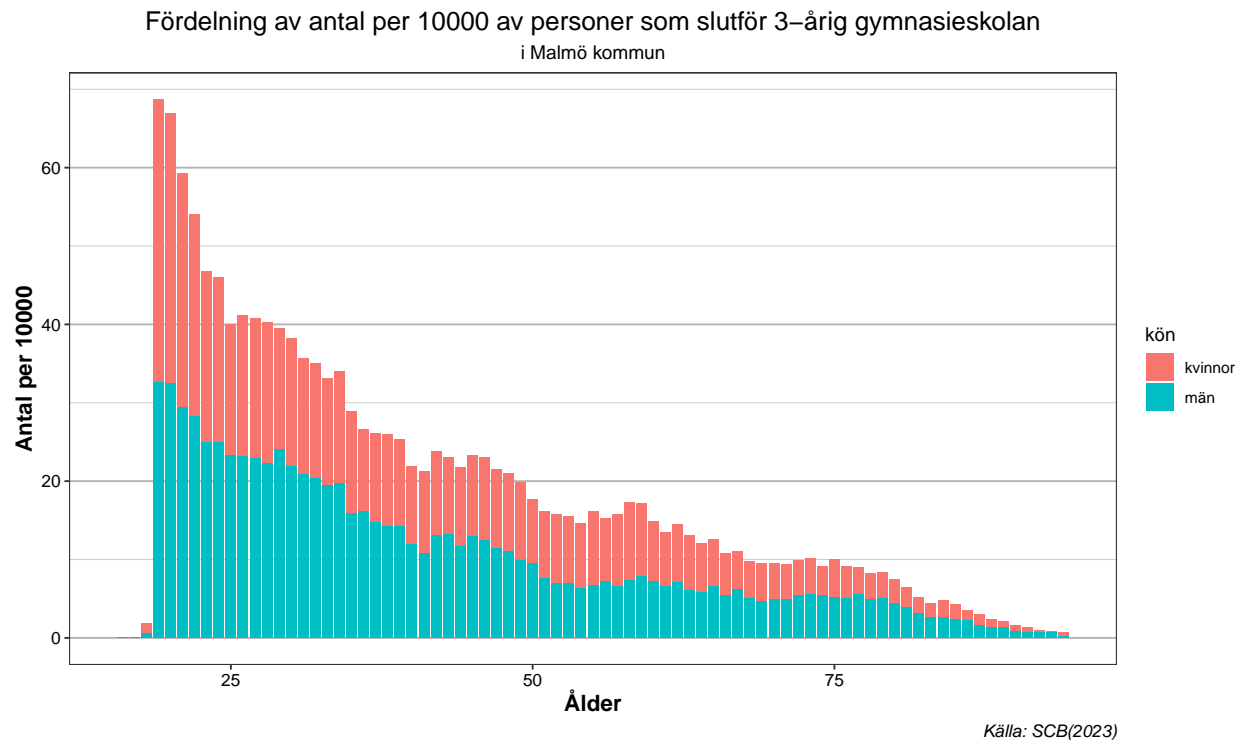
Ovanför visar ett spridningsdiagram likadant med tidigare diagrammet men det är uppdelad efter kön med gröna färg är män och röda färg är kvinnor. Skillnaden är inte stor eftersom både kön visar samma en nedåtgående trend från vänster till höger.

Uppgift 6 (b)



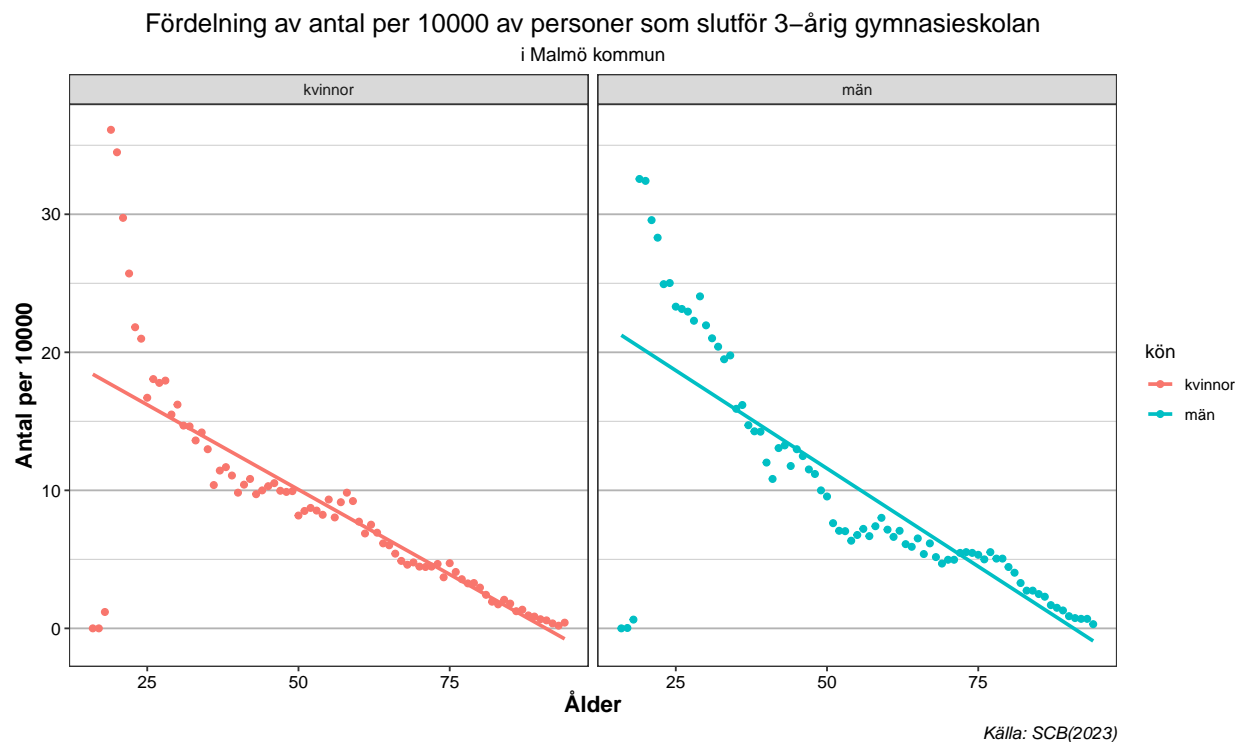
Diagrammet ovan visar ett lådagram med x-axel är ålder och y-axel antal per 10000 och det är också uppdelad efter könen. De flesta observationer av både män och kvinnor har ingen stor skillnad, de ligger mellan 0 till 22 för kvinnor och 25 för män.

Uppgift 6 (c)



Ovanför beskrivs en grupperad stapeldiagram av fördelning av antal personer som slutför 3-årig gymnasieskolan efter ålder på kommunalsnivå med x-axel är ålder och y-axel är antal. Samma med den första diagrammet har det skett ett karftig minskning från vänster till höger, dvs att det spela ingen roll med könen andel av personer som slutför 3-årig gymnasieskolan minskar när ålder ökar.

Uppgift 6 (d)



Samman slutsats med diagrammet 6a men nu det uppdelad till två olika celler istället ligger på en. Både har samma en nedåtgående trend från vänster till höger och skillnaden mellan två könen är inte stor.

Dataanalys av tidseriedata

Uppgift 1

Grafen visar den totala elanvändningen i Sverige från mars 2014 till mars 2024, uppdelad i månader. Elanvändningen i gigawattimmar (GWh) visas på Y-axeln, medan X-axeln visar månaderna under tioårsperioden. Elanvändningen varierar under året och är som högst under de kallaste månaderna (december, januari och februari) på grund av ökat uppvärmningsbehov. Den totala elanvändningen har ökat under de senaste tio åren, med en ökning på cirka 17% från 2014 till 2024.

Uppgift 2

Tabellen och diagrammet visar den genomsnittliga elanvändningen (i GWh) för varje månad. Elanvändningen är som högst under de kallaste månaderna (december, januari och februari) på grund av ökat uppvärmningsbehov, och som lägst under de varmaste månaderna (juli och augusti) på grund av minskat uppvärmningsbehov.

Uppgift 3

Grafen visar den totala elanvändningen i Sverige från 2014 till 2024, uppdelad per år. Elanvändningen i gigawattimmar (GWh) visas på Y-axeln, medan X-axeln visar åren under tioårsperioden. Grafen består av grupperade boxplots som visar den årliga elanvändningen i Sverige. Elanvändningen varierar från år till år, med den högsta användningen 2023 och den lägsta 2014. Det finns en tydlig ökning i elanvändningen från 2014 till 2023.

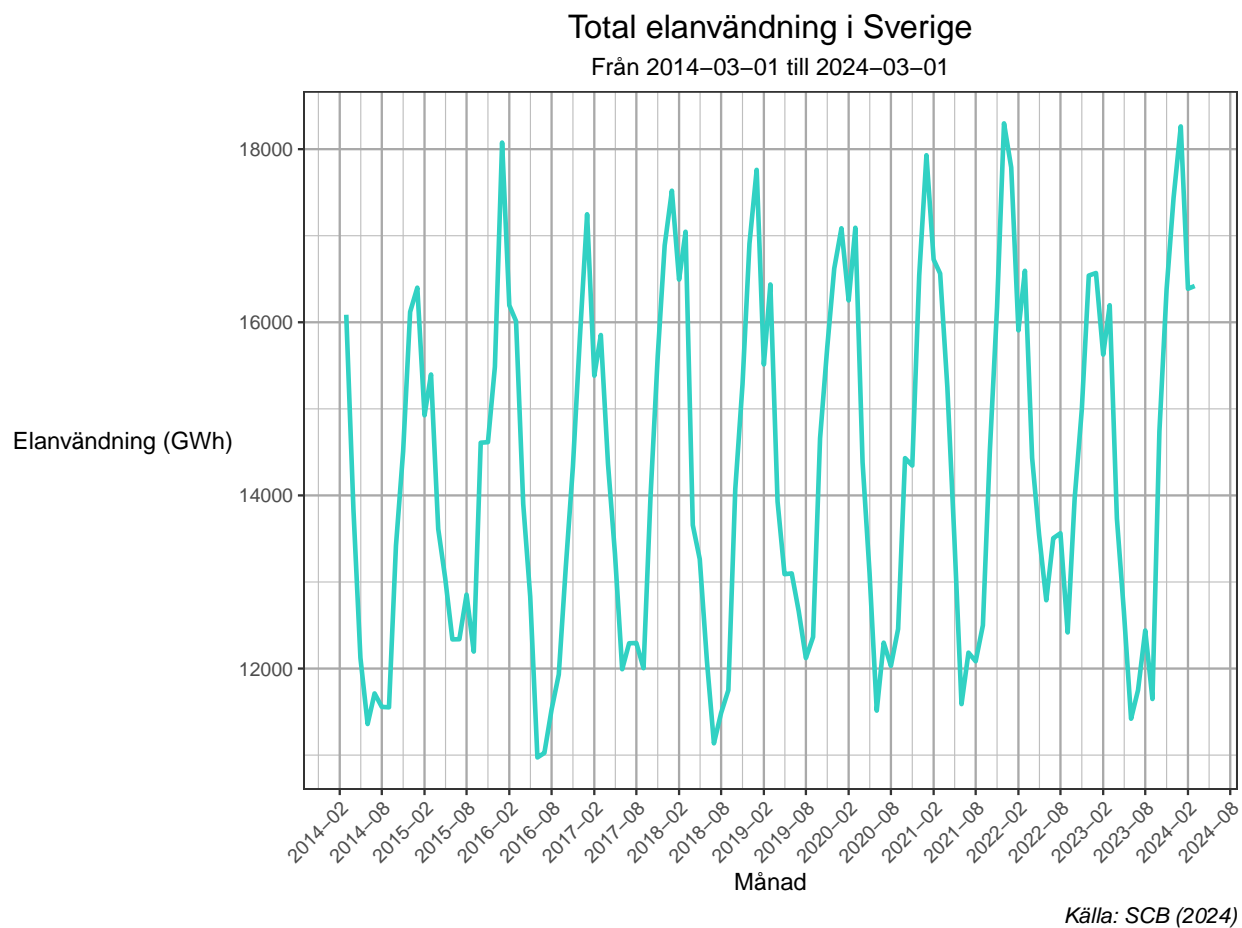


Figure 1: Figur 8: Linjediagram visar total elanvändning i Sverige från 2014-03-01 till 2024-03-01

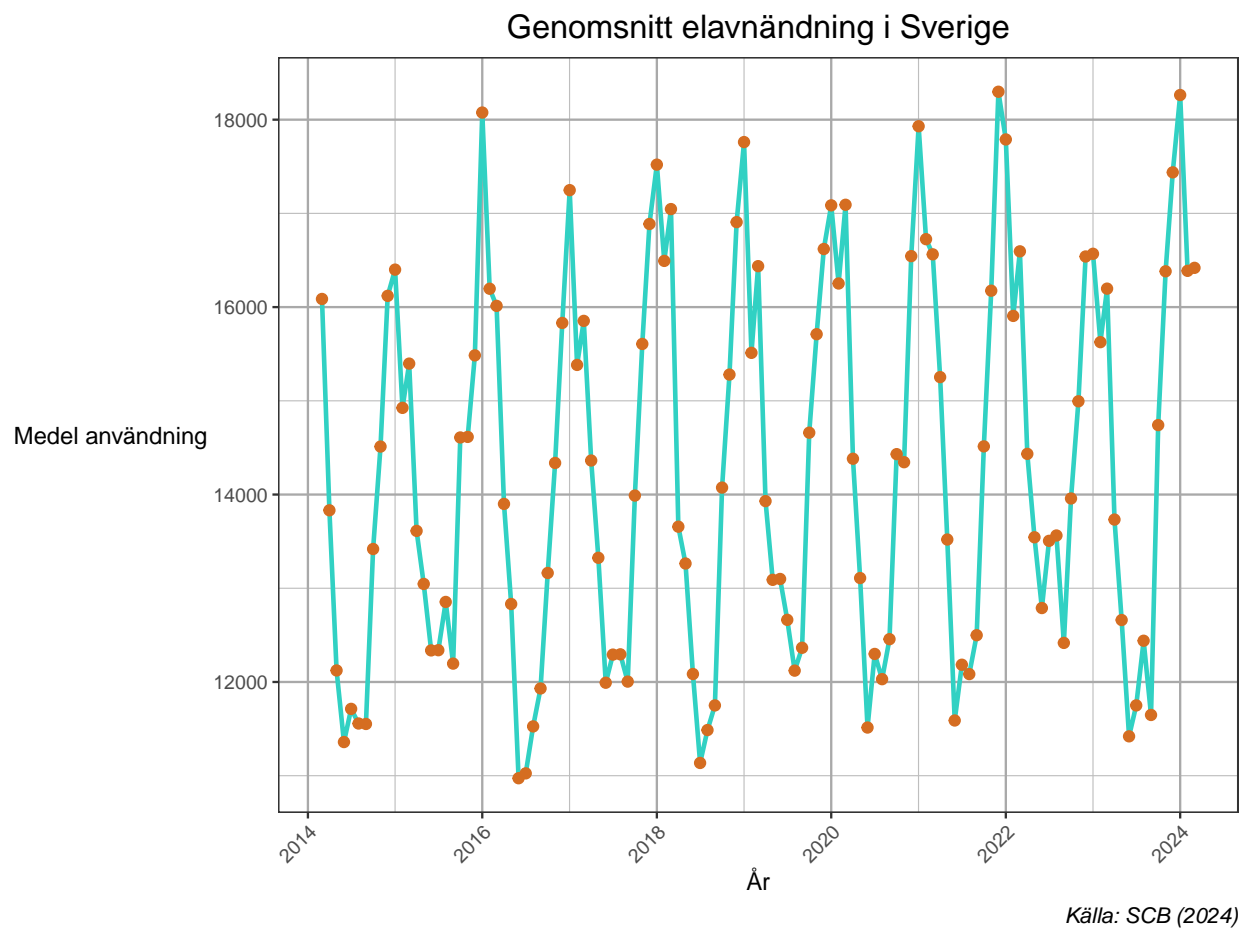


Figure 2: Figur 9: Linjediagram visar genomsnitt elanvändning i Sverige

Table 2: Tabell 2: Genomsnittliga elanvändningen i Sverige från 2014-03-01 till 2024-03-01

månad	mean_elanvandning_gwh
2014-03-01	16087
2014-04-01	13832
2014-05-01	12123
2014-06-01	11359
2014-07-01	11713
2014-08-01	11557
2014-09-01	11552
2014-10-01	13419
2014-11-01	14512
2014-12-01	16120
2015-01-01	16400
2015-02-01	14925
2015-03-01	15397
2015-04-01	13612
2015-05-01	13046
2015-06-01	12337
2015-07-01	12339
2015-08-01	12854
2015-09-01	12196
2015-10-01	14609
2015-11-01	14615
2015-12-01	15485
2016-01-01	18076
2016-02-01	16196
2016-03-01	16013
2016-04-01	13900
2016-05-01	12832
2016-06-01	10973
2016-07-01	11025
2016-08-01	11526
2016-09-01	11931
2016-10-01	13163
2016-11-01	14337
2016-12-01	15831
2017-01-01	17248
2017-02-01	15383
2017-03-01	15853
2017-04-01	14363
2017-05-01	13325
2017-06-01	11992
2017-07-01	12292
2017-08-01	12295
2017-09-01	12004
2017-10-01	13989
2017-11-01	15607
2017-12-01	16887
2018-01-01	17520
2018-02-01	16493
2018-03-01	17046
2018-04-01	13657
2018-05-01	13264
2018-06-01	11
2018-07-01	11136
2018-08-01	11487
2018-09-01	11750

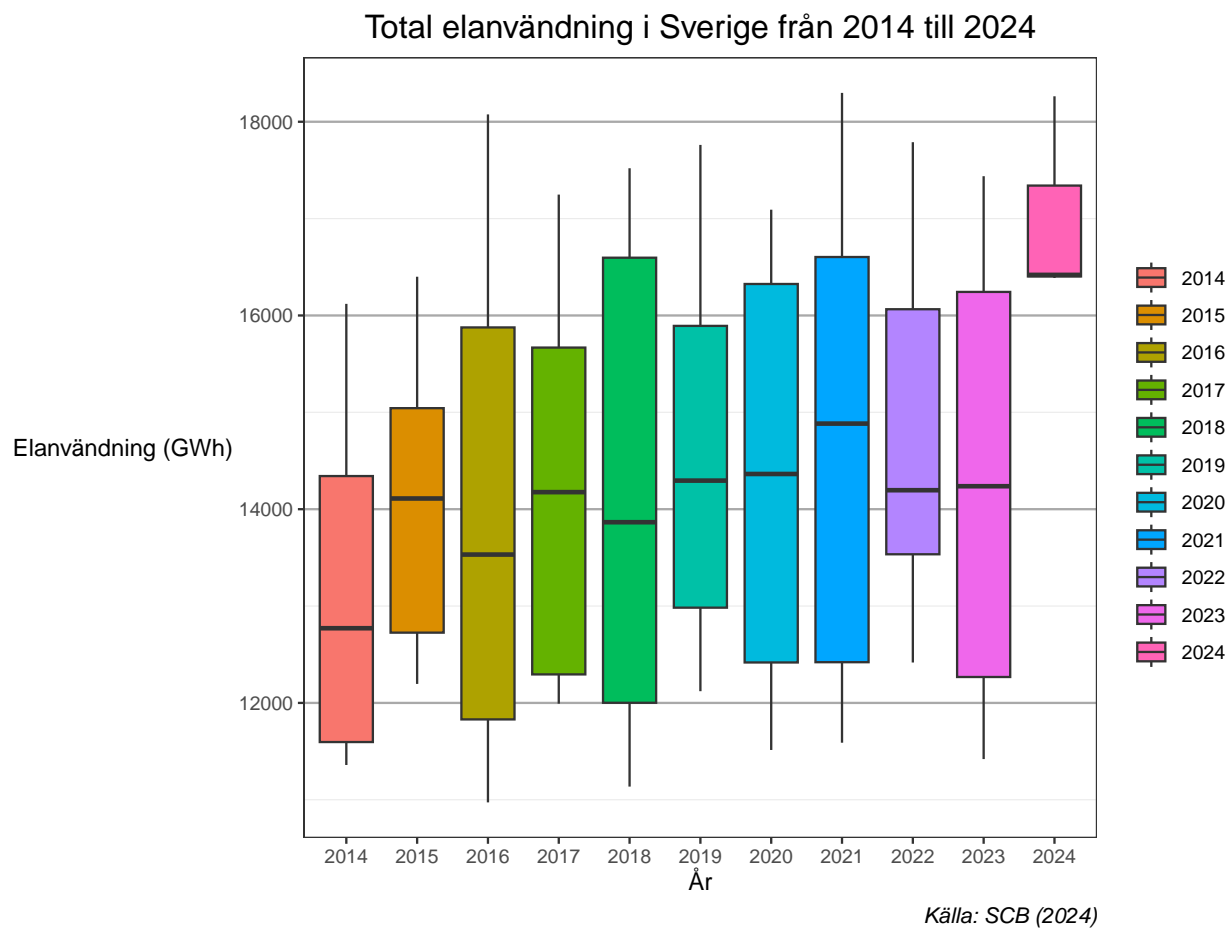


Figure 3: Figur 10: Grupperade boxplots visar den totala elanvändningen i Sverige från 2014 till 2024

Uppgift 4

```
##Uppgift 4##

# Extrahera månaden från datumet
px_data_frame$month <- format(px_data_frame$månad, "%m")

# Beräkna månadsmedelvärden
monthly_averages <- aggregate(Elanvändning_GWh ~ month, data = px_data_frame, FUN = mean)
names(monthly_averages)[2] <- "mean_elanvandning"

# Initiera en ny kolumn för de justerade värdena
px_data_frame$Z <- NA

# Subtrahera månadsmedelvärdet från varje observation
for (i in 1:nrow(px_data_frame)) {
  current_month <- format(px_data_frame$månad[i], "%m")
  monthly_avg <- monthly_averages[monthly_averages$month == current_month, "mean_elanvandning"]
  if (length(monthly_avg) > 0) { # Säkerställ att månadsmedelvärdet finns
    px_data_frame$Z[i] <- px_data_frame$Elanvändning_GWh[i] - monthly_avg
  } else {
    px_data_frame$Z[i] <- NA
  }
}

overall_mean <- mean(px_data_frame$Elanvändning_GWh, na.rm = TRUE)

px_data_frame$Z <- px_data_frame$Z + overall_mean
```

Uppgift 5

Grafen visar "Total elanvändning i Sverige från 2014 till 2024" med justering för säsongsvariation. Den orange linjen representerar den ursprungliga elanvändningen (Y), medan den blå linjen representerar den justerade elanvändningen (Z). Genom att subtrahera den genomsnittliga månatliga elanvändningen har vi tagit bort säsongsvariationen och kunnat se de underliggande trenderna i elanvändningen under tioårsperioden. Den justerade elanvändningen (Z) visar en tydlig ökning av elanvändningen mellan 2014 och 2023.

Uppgift 6

Grafen visar den totala elanvändningen i Sverige från mars 2014 till mars 2024, uppdelad i månader. Elanvändningen i gigawattimmar (GWh) visas på Y-axeln, medan X-axeln visar månaderna under tioårsperioden. Den blå linjen representerar den totala elanvändningen i Sverige under perioden, medan den orange linjen representerar den underliggande trenden i elanvändningen, beräknad med hjälp av en regressionsanalys. Genom att lägga till en regressionslinje till grafen har vi visualiserat den underliggande trenden i elanvändningen i Sverige under perioden 2014 till 2024. Trenden visar en tydlig ökning i elanvändningen under denna period.

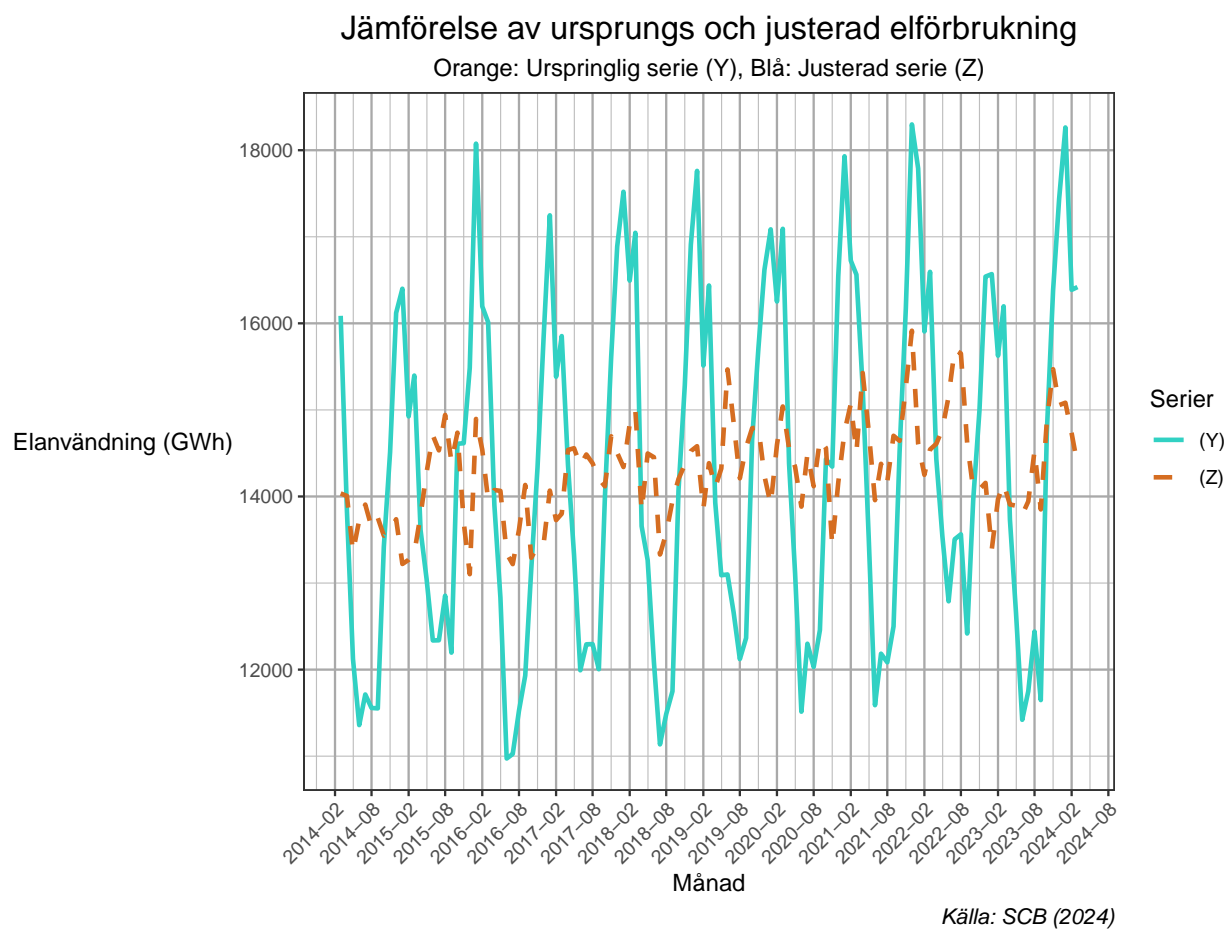


Figure 4: Figur 11: Linjediagram visar total elanvändning i Sverige från 2014 till 2024 med justering för säsongsvariation

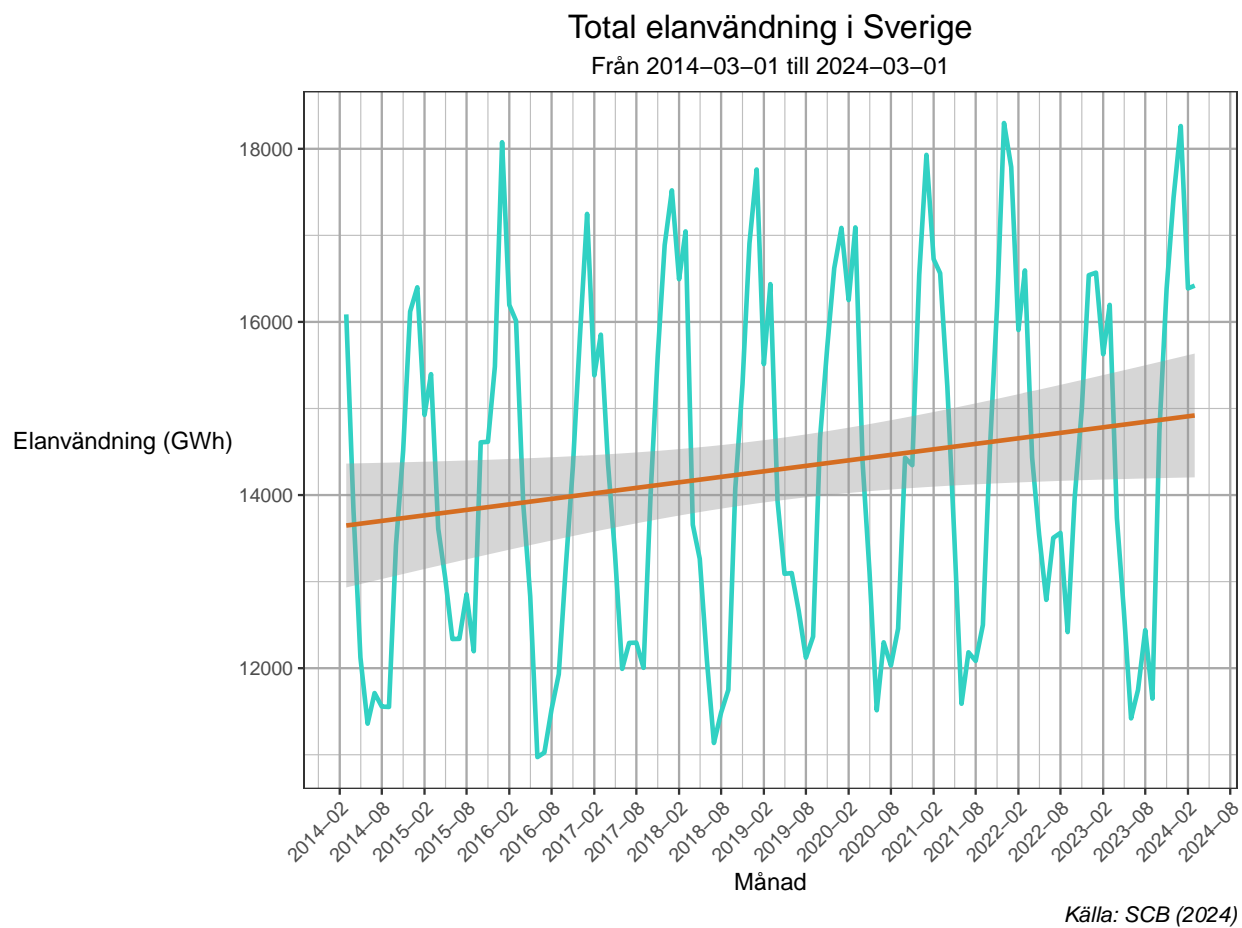


Figure 5: Figur 12: Linjediagram med n regressionslinje visar den totala elanvändningen i Sverige från mars 2014 till mars 2024

Uppgift 7

Från den regressionslinje som lagts till i tidsseriediagrammet (se Graf 12) framgår det att det finns en något ökande trend i elanvändningen under perioden mars 2014 till mars 2024. Den orange regressionslinjen, även om den inte är brant, tyder på en gradvis ökning i den totala elförbrukningen. Det månatliga medelvärdet (se Graf 9) och boxplots per år (se Graf 10) indikerar signifikant säsongsvariation i data. Till exempel:

- Elanvändningen tenderar att vara högre under vintermånaderna (december till februari) och lägre under sommarmånaderna (juni till augusti).
- Detta mönster är konsekvent över åren, vilket indikerar att säsongsmässiga förändringar (t.ex. ökad uppvärmning under vintern och minskad användning under sommaren) avsevärt påverkar elförbrukningen.

Sammanfattningsvis visar data en svagt ökande trend i elförbrukningen över åren, tillsammans med ett tydligt säsongsmönster, med högre användning på vintern och lägre användning på sommaren.