

Inlämningsuppgift 1

Grupp 19, Erik Borén

Innehållsförteckning

1	Inlämningsuppgift 1	1
1.1	Uppgift 1	1
1.2	Uppgift 2	2
1.3	Uppgift 3	3

1 Inlämningsuppgift 1

1.1 Uppgift 1

a) Vilken information ger visualiseringen?

Visualiseringen kommunicerar orsaker till förtida död som andel av total antal år “förlorade” av förtida död. Utöver detta så används olika färger för att dela in dödsorsaker i tre olika kategorier samt används en färgskala för varje kategori som visar förändringen per år mellan 2005 och 2010 där en mörkare färg indikerar en ökning, på samma sätt indikerar en ljusare nyans att det skett en minskning.

b)

- Storlek: År förlorade till följd av förtidada död som en andel av totalt antal förlorade år.
- Färg: Kategorin som dödsorsaken tillhör.
- Färgton: Förändring per år mellan 2005 och 2010.

c) Hur påverkar 3D-effekten möjligheten att jämföra areornas storlekar?

3D-effekten försvårar uppskattning av relativ storlek mellan olika sektorer i visualiseringen. Detta på grund av den introducerar ett *djup* i bilden.

d) Beskriv ytterligare styrkor/svagheter med visualiseringen utifrån olika principer om visualisering.

Heer & Bostock (2010) kommer fram till resultatet att area är bland det svåraste att uppskatta korrekt. Valet av typ av visualisering är både en styrka och svaghet och behöver sättas emot vad syftet för visualiseringen. 3D-effekten är visserligen snygg och gör att visualiseringen fånga uppmärksamhet men den hjälper inte till att tyda diagrammet eller att förmedla slutsatser.

Att visualiseringen visar år förlorade till följd av förtida död är borde uppmärksammas tydligare. "Total life-years lost" nämns enbart i en två små texter kopplade till två areor.

e) Kan du föreslå några förbättringar? Hur skulle detta kunna ha gjorts bättre?

- Förtydliga att visualiseringen visar life years lost.
- Inkludera siffror för att uppskatta storleken av vissa kategorier som är av intresse.
- Använd en 2D-visualisering istället för 3D.

1.2 Uppgift 2

```
library(tidyr)
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

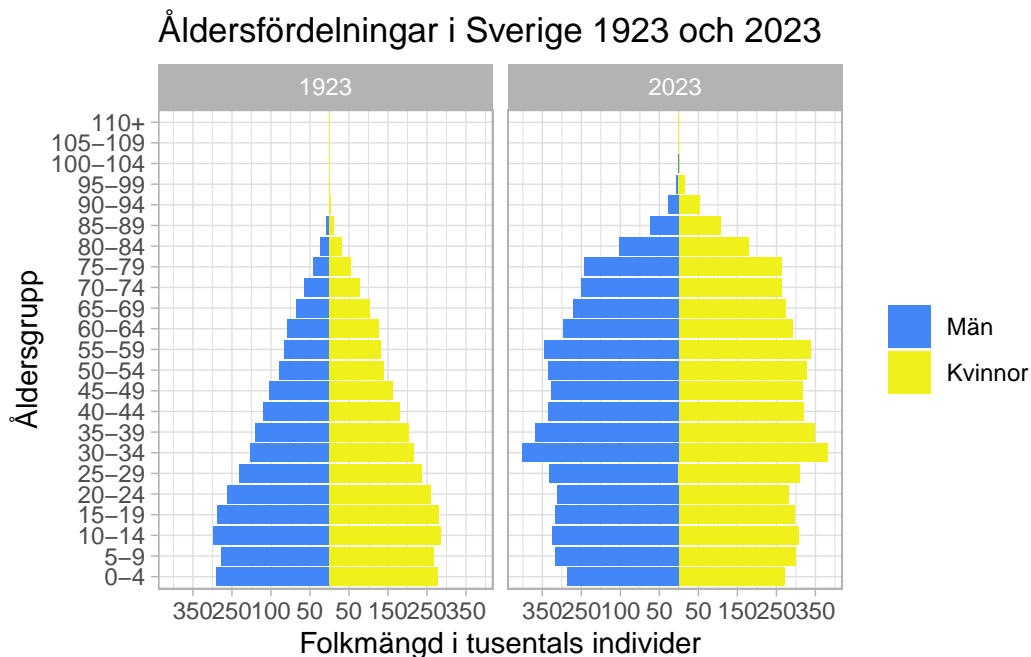
The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
library(ggplot2)
```

```
load(file = "inlämningsuppgift_1.rda")
pyramid <- ggplot(data = population_1923_2023,
  aes(x = count,
      y= reorder(age, order),
      fill = sex))+
  geom_col(data = subset(population_1923_2023, sex=="men"),
    aes(x = -count)) +
  geom_col(data = subset(population_1923_2023, sex=="women"),
    aes(x = count)) +
  scale_x_continuous(breaks = c(-350000, -250000, -150000, -50000, 50000, 150000, 250000, 350000),
    labels = c(350, 250, 100, 50, 50, 150, 250, 350))+
  facet_grid(cols= vars(year)) +
  labs(title = "Åldersfördelningar i Sverige 1923 och 2023", x = "Folkmängd i tusentals individer")
  theme_light() +
  theme(legend.title=element_blank())+
  scale_fill_manual(values = c("#4287f5", "#f0f01a"), labels = c("Män", "Kvinnor"))

pyramid
```



1.3 Uppgift 3

Jag har valt att visa varje åldersgrupps andel av den totala befolkningen för varje år.

Detta gör jag genom att först beräkna varje åldersgrupps andel av den totala befolkningen för varje år och sedan presenterar jag detta i ett stolpdiagram. Eftersom att jag vill att den yngsta åldersgruppen skall vara längst ned behövde jag lägga in en data frame med order och de åldersgrupper jag har och sedan utföra en join med underlaget från SCB. På så sätt kan jag få den ordning som eftersträvas.

```
age_order <- data.frame(
  order = c(1:10),
  age = c("0-9", "10-19", "20-29", "30-39", "40-49", "50-59", "60-69", "70-79", "80-89", "90+")
)

upp3data <-
  population_1860_2023 %>%
  group_by(year, age) %>%
  drop_na() %>%
  summarize(count = sum(count)) %>%
  mutate(freq = count / sum(count)) %>%
  inner_join(age_order)
```

`summarise()` has grouped output by 'year'. You can override using the
 `groups` argument.
 Joining with `by = join_by(age)`

```
upp3data
```

```
# A tibble: 1,640 x 5
# Groups:   year [164]
   year age    count    freq order
  <chr> <chr>  <dbl>   <dbl> <int>
1 1860 0-9    921411 0.239     1
2 1860 10-19  725609 0.188     2
3 1860 20-29  619788 0.161     3
4 1860 30-39  568060 0.147     4
5 1860 40-49  421105 0.109     5
6 1860 50-59  286178 0.0741    6
7 1860 60-69  209108 0.0542    7
8 1860 70-79   88602 0.0230    8
9 1860 80-89   18478 0.00479   9
10 1860 90+     1389 0.000360 10
# i 1,630 more rows
```

```

upp3 <- ggplot(data = upp3data,
               aes(x=as.numeric(year),
                   y=freq,
                   fill = reorder(age,-order)))+
  geom_col()+
  labs(title = "Åldersfördelning i Sverige 1860 - 2023",
       x = "År",
       y = "Procent av befolkning",
       fill = "Åldersgrupp")+
  theme_minimal()+
  scale_y_continuous(labels = scales::percent)

```

upp3

