worksheet 9

2024-07-24

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
hibbs <- as_tibble(read.csv("/Users/ella/Downloads/hibbs.dat",sep=""))
```

Problem 1

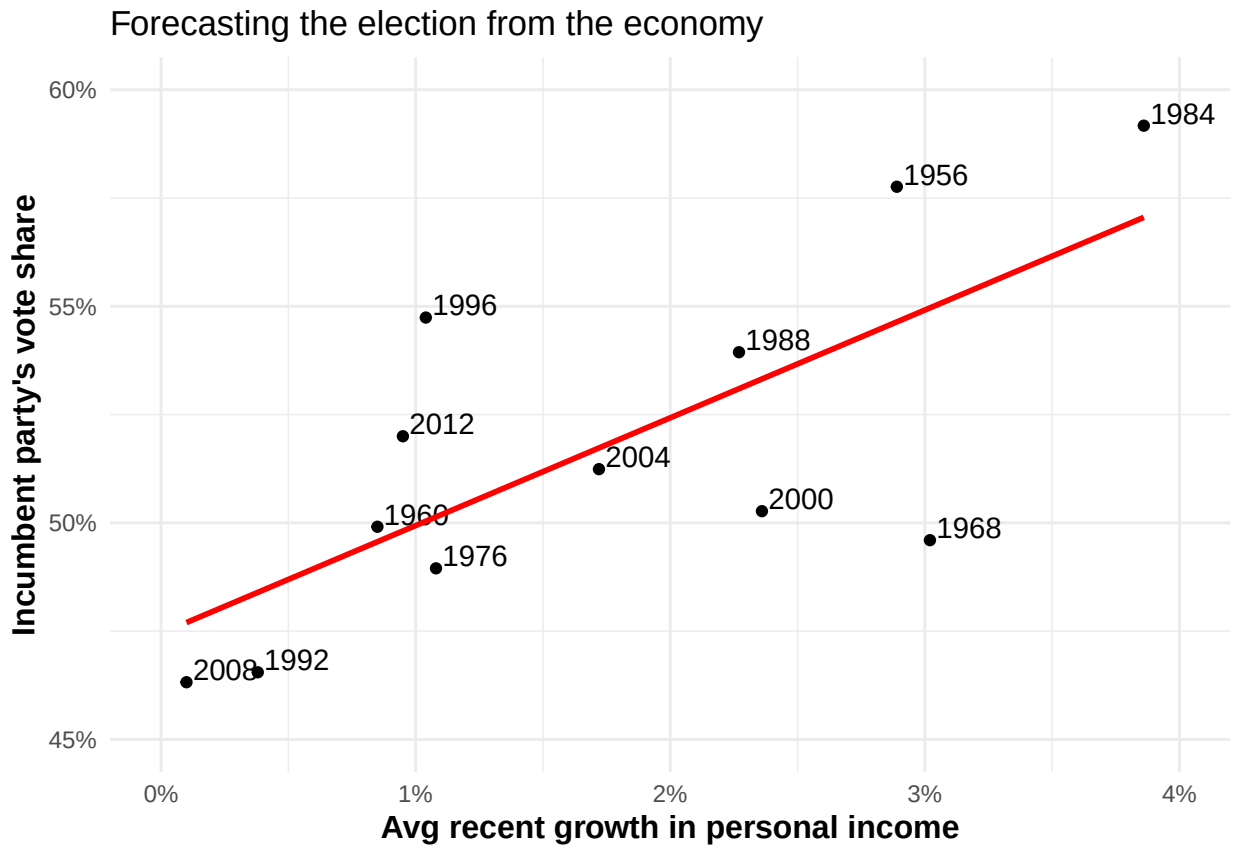
```
library(ggplot2)
ggplot(hibbs, aes(x = growth, y = vote)) +
  geom_point() +
  geom_text(aes(label = year), hjust = -0.1, vjust = -0.1) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Forecasting the election from the economy",
       x = "Avg recent growth in personal income",
       y = "Incumbent party's vote share") +
  scale_x_continuous(labels = scales::percent_format(scale = 1), limits = c(0, 4), breaks = seq(0, 4, by
  scale_y_continuous(labels = scales::percent_format(scale = 1), limits = c(45, 60), breaks = seq(45, 60
  theme_minimal() +
  theme(
    axis.title.x = element_text(size = 12, face = "bold"),
    axis.title.y = element_text(size = 12, face = "bold")
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 4 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

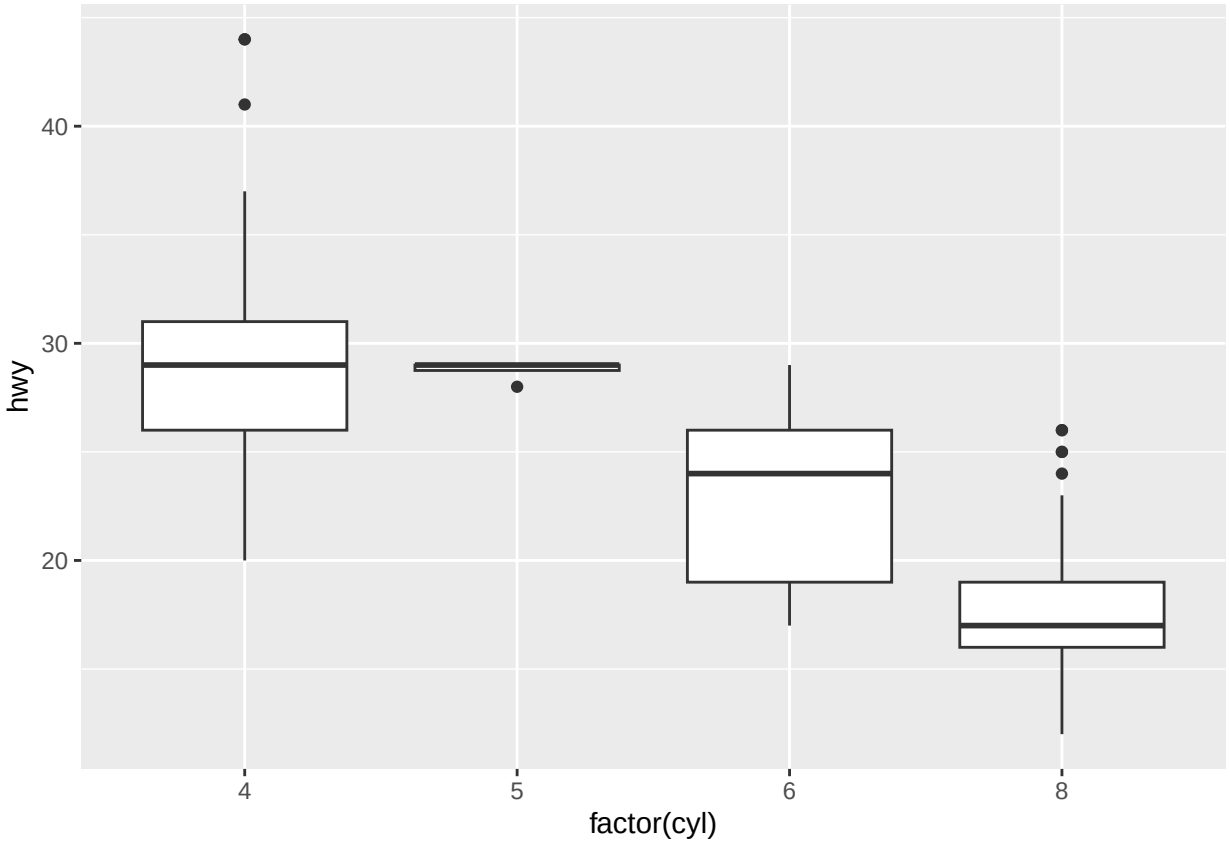
```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_text()`).
```



Problem 2

```
?mpg
p <- ggplot(mpg, mapping = aes(x = factor(cyl), y = hwy))
p + geom_boxplot()
```

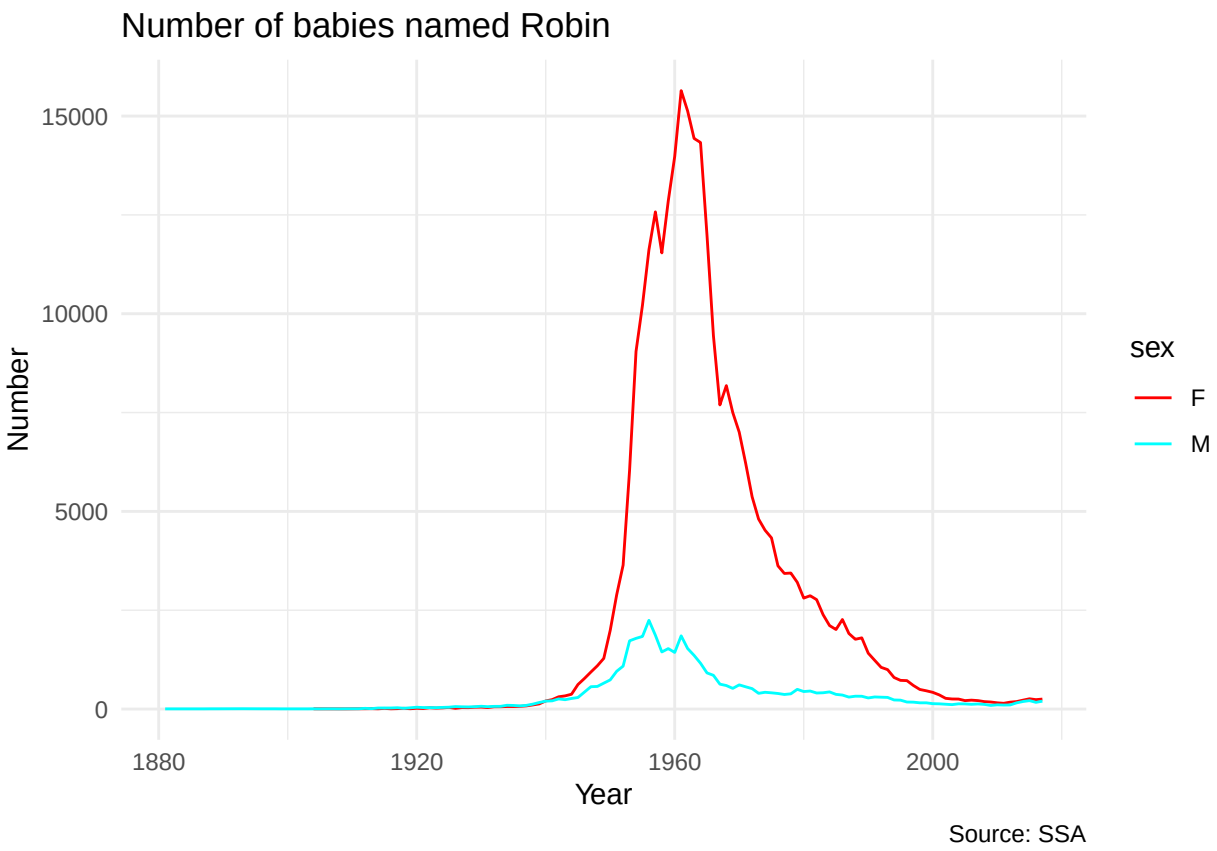


Problem 3

```
library(babynames)
robin <- filter(babynames, name == "Robin")
head(robin, 4)
```

```
## # A tibble: 4 x 5
##    year sex   name      n      prop
##   <dbl> <chr> <chr> <int>     <dbl>
## 1  1881 M     Robin     5 0.0000462
## 2  1887 M     Robin     5 0.0000457
## 3  1888 M     Robin     6 0.0000462
## 4  1889 M     Robin     6 0.0000504
```

```
ggplot(robin, aes(x = year, y = n, color = sex)) +
  geom_line() +
  labs(title = "Number of babies named Robin",
       x = "Year",
       y = "Number",
       caption = "Source: SSA") +
  theme_minimal() +
  scale_color_manual(values = c("F" = "red", "M" = "cyan"))
```


Number of babies named Robin
15000
10000
5000
0
Number
1880
1920
1960
2000
Year
sex
F
M
Source: SSA