

Homework 1

Zahlen Zbinden

2024-01-15

```
font_add_google("Poppins", "pop")
txt_font = "pop"
showtext_auto()
```

```
retail <- read_excel("D:/RepoMan/osu/data/NewHomeSales.xls", sheet = "CIDR")
```

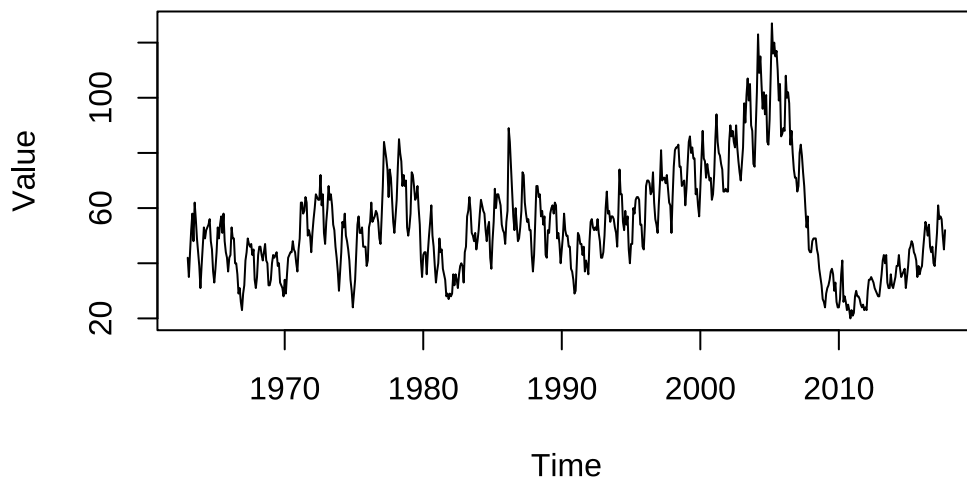
```
retail <- retail |>
  mutate(Value = as.numeric(Value)) |>
  na.omit() |>
  mutate(Period = my(Period)) |>
  mutate(Period = as.Date(Period, "%b-%Y"))
```

```
print(retail[1:24,])
```

```
# A tibble: 24 x 2
  Period      Value
  <date>      <dbl>
1 1963-01-01    42
2 1963-02-01    35
3 1963-03-01    44
4 1963-04-01    52
5 1963-05-01    58
6 1963-06-01    48
7 1963-07-01    62
8 1963-08-01    56
9 1963-09-01    49
10 1963-10-01    44
# i 14 more rows
```

```
ts_data <- ts(
  retail[,2],
  start = c(1963, 1),
  frequency = 12
)

plot(ts_data)
```



```
retail |>
  ggplot(
    aes(x = Period, y = Value)
  ) +
    geom_area(
      fill = "firebrick",
      alpha = .5
    ) +
    geom_point(
      size = 2
    ) +
    geom_line(
```

```

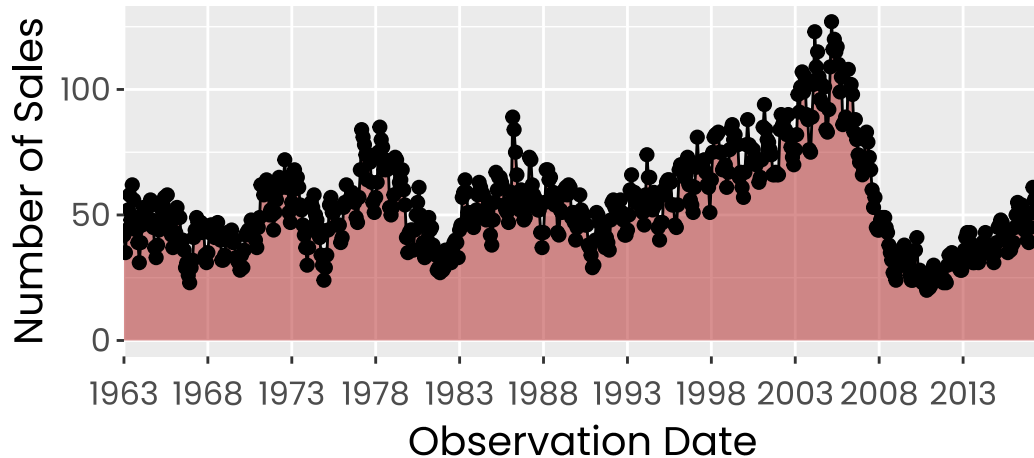
    size = .5
  ) +
  scale_x_date(
    limits = as.Date(c("1963-01-01", "2017-09-01")),
    breaks = "5 years",
    date_labels = "%Y",
    expand = c(0, 0)
  ) +
  labs(
    title = "New Home Sales",
    subtitle = "Monthly Retail Trade Survey",
    x = "Observation Date",
    y = "Number of Sales"
  ) +
  theme(
    text = element_text(
      family = txt_font,
      size = 15
    ),
    plot.title = element_text(
      size = 40,
    ),
    plot.subtitle = element_text(
      size = 20
    ),
    axis.title.x = element_text(
      vjust = -1.5,
      margin = margin(0,0,10, 0)
    ),
    panel.grid.minor.x = element_blank(),
    axis.text.x = element_text(
      vjust = -.5
    )
  )
)

```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
 i Please use `linewidth` instead.

New Home Sales

Monthly Retail Trade Survey



1d. The variance does not change throughout time, as we see the same ups and downs throughout the observations. We can see there is a seasonality in new home sales that ramps up and down throughout each year of observations. There are not any outliers that can be seen from the plot. The series increases from 1963 to approximately 2005 when we can see a sharp decline this marks the beginning of the market crash.

```
hare <- read_excel("D:/RepoMan/osu/data/hare.xlsx", sheet = "hare")
```

```
sapply(hare, class)
```

```
Year Abundance  
"numeric" "numeric"
```

```
hare <- hare |>  
  mutate(Abundance = as.numeric(Abundance)) |>  
  mutate(Year = as.numeric(Year)) |>  
  mutate(Year = ymd(Year, truncated = 2L))
```

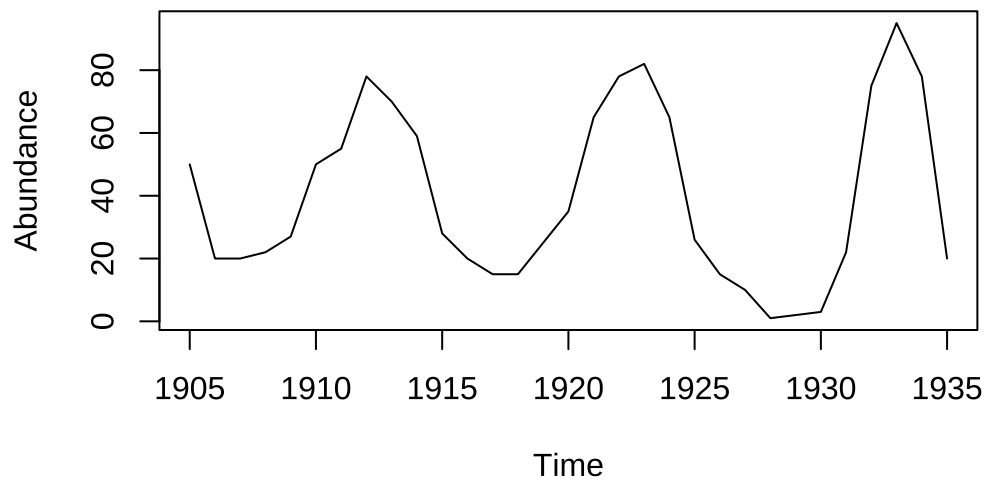
```
ts_data <- ts(  
  hare[,2],
```

```

      start = 1905,
      end = 1935,
      frequency = 1
    )

```

```
plot(ts_data)
```



```
hare <- read_excel("D:/RepoMan/osu/data/hare.xlsx", sheet = "hare")
```

```

hare <- hare |>
  mutate(Abundance = as.numeric(Abundance)) |>
  mutate(Year = as.numeric(Year))

```

```

hare |>
  ggplot(
    aes(x = Year, y = Abundance)
  ) +
  geom_point(
    shape = 1,
    size = 3
  )

```

```

) +
geom_line() +
scale_x_continuous(
  breaks = seq(
    from = 1905,
    to = 1935,
    by = 5
  )
) +
scale_y_continuous(
  breaks = seq(
    from = 0,
    to = 80,
    by = 20
  )
) +
theme(
  text = element_text(
    family = txt_font,
    size = 15
  ),
  plot.title = element_text(
    size = 40,
  ),
  plot.subtitle = element_text(
    size = 20
  ),
  axis.title.x = element_text(
    vjust = -1.5,
    margin = margin(0,0,10, 0)
  ),
  axis.title.y = element_text(
    vjust = 4,
    margin = margin(0,0,0,20)
  ),
  axis.text.x = element_text(
    vjust = -.5
  ),
  axis.text.y = element_text(
    hjust = -.5
  ),

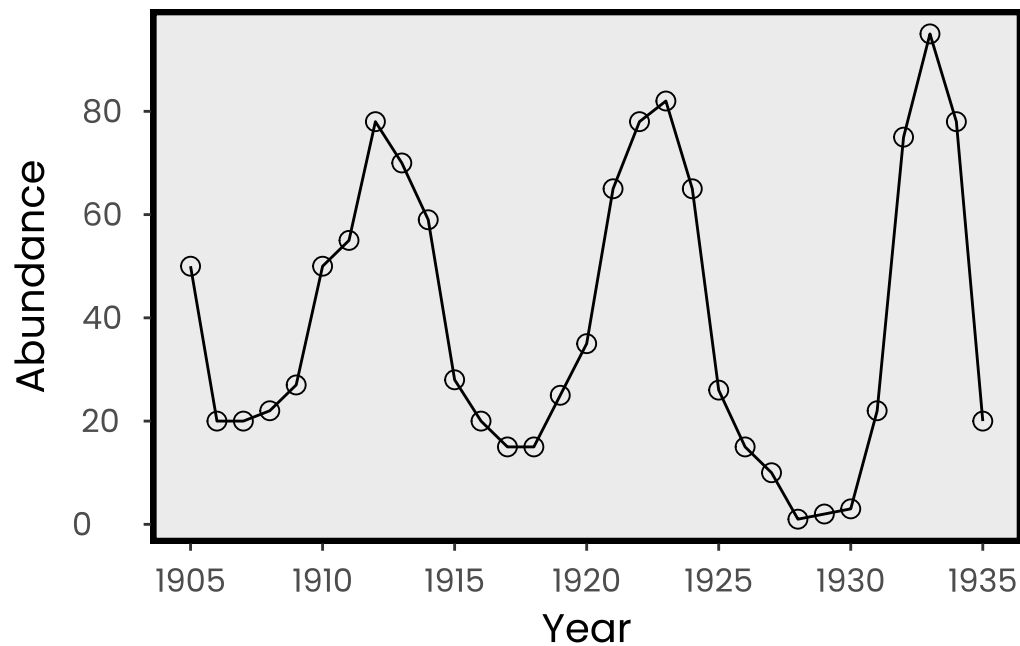
```

```

panel.grid = element_blank(),
panel.border = element_rect(
  size = 2,
  fill = NA
)
)

```

Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
 i Please use the `linewidth` argument instead.

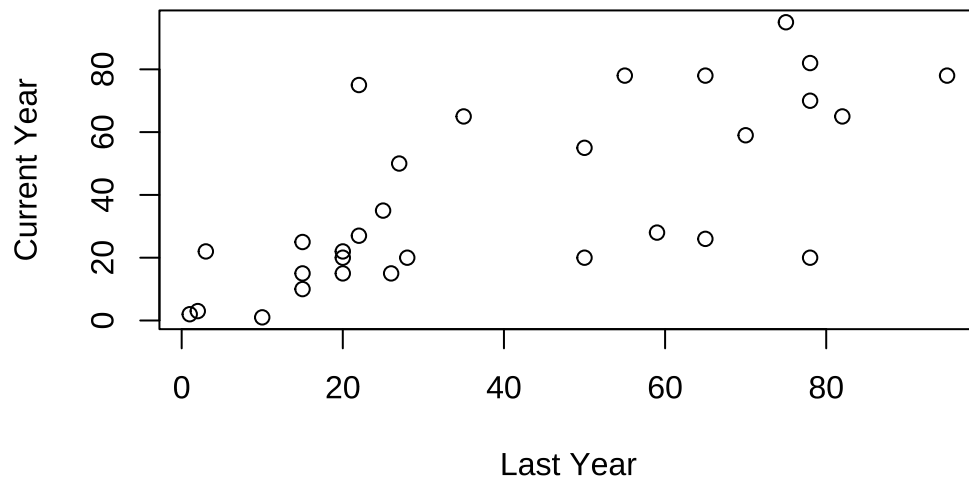


1c. The data seems to show some sort of cyclic variation. There is some time period that may be predictable. We can see that there is a rise and fall that seem to occur in intervals. There doesn't seem to be a trend in the changing of the mean as the high points get higher so do the low points.

```

lag1 <- zlag(ts_data, d = 1)
plot(lag1, ts_data, xlab = "Last Year", ylab = "Current Year")

```



1d. From the scatter plot of previous year to current year it does appear that there is an increase in the total number of hairs per year, as we can see that current year to last year both increase, we can also observe this from the time series plot as the spread increases dramatically.