Solar Cyles

Team Blue

3/30/2022

Loading libraries

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
library(lubridate)
library(timetk)
```

```
## Warning: package 'timetk' was built under R version 4.1.3
```

1

```
sunspots <- read_delim(
  "SN_d_tot_V2.0.csv",
  delim = ";",
  escape_double = FALSE,
  col_names = c(
    "year",
    "month",
    "day",
    "year_fraction",
    "ss_number",
    "sd",
    "n_obs",
    "definitive"
  ),
  col_types = "iiininii",
  na = c("-1", "-1.0"),
  trim_ws = TRUE
)
```

The variables are year, month, day, date in fraction of year, daily total sunspot number, daily standard deviation of the sunspot numbers, number of observations, and definitive/provisional indicator, respectively.

Missing values are "-1" or "-1.0".

2

```
sunspots <- mutate(sunspots, date = make_date(year, month, day))
```

3

```
my_stamp <- function(x) {
  stamp("Tuesday, 3 December, 2019", quiet = TRUE)(x)
}

sunspots |>
  filter(is.na(ss_number)) |>
  slice_max(n = 1, date) |>
  pull(date) |>
  my_stamp()
```

```
## [1] "Friday, 22 December, 1848"
```

4

```
sunspots <-
  sunspots |>
  mutate(diff_to_prev_data = date - lag(date))

# Checking if all data points have a difference of 1 day from the previous data.
all(sunspots$diff_to_prev_data == 1, na.rm = TRUE)
```

```
## [1] TRUE
```

Yes, they are all consecutive.

5

```
spots <-
  sunspots |>
  summarise_by_time(
    date,
    .by = "month",
    mean_daily_ss_number = mean(ss_number, na.rm = TRUE)
  ) |>
  filter(date >= ymd("1975-01-01"))
```

6

```
solar_cycles <-
  read_csv("solar_cycles.csv") |>
  mutate(
    `Start (Minimum)` = ym(`Start (Minimum)`),
    Maximum = ym(Maximum)
  )
```

7

```
### SOLVE###
solar_cycles <-
  solar_cycles |>
  mutate(
    isequal = round(
      interval(
        `Start (Minimum)`, Maximum
      ) / years(1),
      digits = 1
    ) == `Time of Rise (years)`
  )
table(solar_cycles$isequal)
```

```
##
## FALSE  TRUE
##     2    22
```

All are equal except two entries.

8

```
solar_cycles <- filter(solar_cycles, `Start (Minimum)` >= ymd("1975-01-01"))
```

9

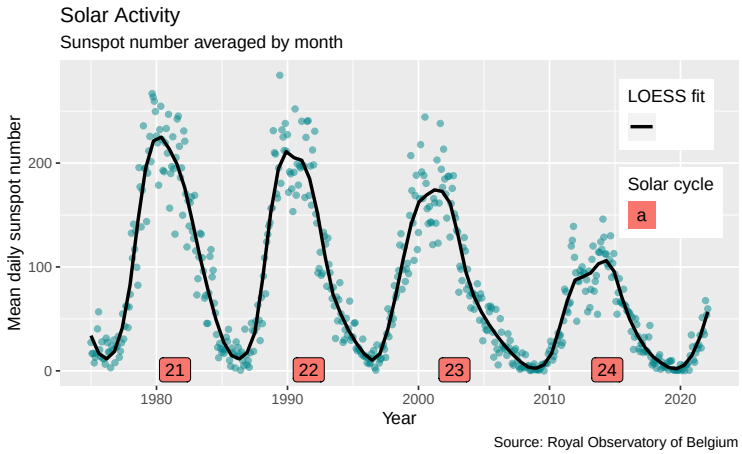
```
solar_cycles <-
  solar_cycles |>
  mutate(
    next_start = lead(`Start (Minimum)`),
    mid_date = `Start (Minimum)` + ((next_start - `Start (Minimum)`) / 2),
    next_start = NULL
  )
```

10

```
solar_cycles <-
  solar_cycles |>
  mutate(`Solar Cycle Number` = 21:25)

ggplot(data = spots, aes(date, mean_daily_ss_number)) +
  geom_point(alpha = 0.5, color = "turquoise4") +
  labs(
    title = "Solar Activity",
    subtitle = "Sunspot number averaged by month",
    caption = "Source: Royal Observatory of Belgium",
    x = "Year",
    y = "Mean daily sunspot number",
    linetype = "LOESS fit",
    label = "Solar cycle",
    fill = "Solar cycle"
  ) +
  geom_smooth(se = FALSE, span = 0.1, color = "black", aes(linetype = "")) +
  geom_label(
    data = solar_cycles,
    aes(mid_date, 1, label = `Solar Cycle Number`, fill = "")
  ) +
  theme(legend.position = c(0.9, 0.70))
```

```
## Warning: Removed 1 rows containing missing values (geom_label).
```



11

The plot reveals that sunspot density generally happens in cycles. We see that there is a trough, whereby there is generally lower sunspot density around the years of 1976, 1986, 1997, 2009, 2019. We also see there is a peak, whereby there is generally higher sunspot density around the years of 1980, 1990, 2002, 2014. This is plausible, as solar eclipses and other solar occurrences also happen on a cyclical basis. What is interesting is that the peaks are seen to have a decreasing trend over the past 40 years and I speculate that this may be due to climate change. Increased CO2 in the atmosphere may reduce the number of sunspots visible on the sun.