

Temperature

Denisa Mensatorisova

AIR-TEMPERATURE-OBSERVATION air temperature

The temperature of the air.

MIN: -0932 MAX: +0618 UNITS: Degrees Celsius

SCALING FACTOR: 10

According to isd documentation, under TMP field there are two information. AIR-TEMPERATURE-OBSERVATION air temperature which concerns us and should be numeric, with sign prefix. If any value is missing 9999 represents it. It is scaled up tenfold.

Second information does not concern us and is thus removed. Both values are separated and only temperature is kept. Temperature is then converted to numeric (if not 9999) and divided by 10.

Teplota vzduchu meraná v °C.

```
all_data <- read.csv(file= "../data/all.csv")
head(all_data)

##      X      STATION          DATE   TMP WND_ANGLE WND_SPEED   VIS SLP    CIG DEW
## 1 1 11927599999 2004-05-10T00:00:00     8       NA         1 11265  NA 3000    8
## 2 2 11927599999 2004-05-10T01:00:00     8       NA         0 11265  NA 3000    8
## 3 3 11927599999 2004-05-10T02:00:00     7      330         1 11265  NA 22000   7
## 4 4 11927599999 2004-05-10T03:00:00     6       NA         0 11265  NA 22000   6
## 5 5 11927599999 2004-05-10T04:00:00     7       NA         1 1500   NA 120    7
## 6 6 11927599999 2004-05-10T05:00:00     7       NA         1 3500   NA 120    7

##      SNOW_DEPTH PWO GSO    LP LP24
## 1           NA  NA  NA <NA>   NA
## 2           NA  NA  NA <NA>   NA
## 3           NA  NA  NA <NA>   NA
## 4           NA  NA  NA <NA>   NA
## 5           NA  NA  NA <NA>   NA
## 6           NA  NA  NA <NA>   NA

summary(all_data$TMP)

##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.   NA's
## -28.700   2.000  9.200  9.482 16.500  38.000    341

all_data[ 'TMP' ] %>% profiling_num()

##      variable      mean std_dev variation_coef p_01 p_05 p_25 p_50 p_75 p_95 p_99
## 1      TMP 9.481857 9.571159     1.009418 -12 -5.4     2 9.2 16.5 25.1   30
##      skewness kurtosis iqr range_98 range_80
## 1 -0.00843519 2.591749 14.5 [-12, 30] [-2.3, 22]

getmode(na.omit(all_data$TMP)) # modulus

## [1] 1
```

Centrálna poloha dát

Hodnota výberového mediánu je 9.2 a hodnota výberového priemeru 9.482. Rozdiel hodnôt je iba 0.3 teda výberový priemer môžeme považovať za dobrý odhad strednej hodnoty.

Modus - najčastejšia hodnota teploty je 1.

```
var(all_data$TMP, na.rm = T) # rozptyl  
  
## [1] 91.60708  
max(all_data$TMP, na.rm = T) - min(all_data$TMP, na.rm = T) # variacne rozpatie  
  
## [1] 66.7  
# Interquartile range and outliers  
Q1 <- quantile(all_data$TMP, 0.25, na.rm = T) # 25% hodnot je mensich a 75% vacsich  
Q3 <- quantile(all_data$TMP, 0.75, na.rm = T) # 75% hodnot je mensich a 25% vacsich  
IQR <- IQR(all_data$TMP, na.rm = T) # interquartile range  
IQR_dev <- IQR/2  
  
# odlahlé hodnoty  
length(which(all_data$TMP < (Q1 - 1.5*IQR)))  
  
## [1] 403  
length(which(all_data$TMP > (Q3 + 1.5*IQR)))  
  
## [1] 0  
# extremne hodnoty  
length(which(all_data$TMP < (Q1 - 3*IQR)))  
  
## [1] 0  
length(which(all_data$TMP > (Q3 + 3*IQR)))  
  
## [1] 0
```

Variabilita

Výberový rozptyl je 91.607, čo je stredná kvadratická odchýlka hodnôt od výberového priemeru. Vyjadruje ako veľmi sú hodnoty rozptýlené od priemeru. Čím je väčší, tým viac sa hodnoty odchylujú od priemeru.

Výberová smerodajná odchýlka je 9.5712. Vyjadruje kvadratický priemer odchýlok jednotlivých hodnôt teploty od ich aritmetického priemeru. Jej hodnota môže byť vyššia v porovnaní s medzikvartilovou odchýlkou pretože v dátach sa nachádza niekoľko odlahlých hodnôt.

Medzikvartilová odchýlka (IQR/2) je 7.25, čo je v porovnaní so smerodajnou odchýlkou o niečo menej.

Variačné rozpätie je 66.7. Daná hodnota predstavuje rozdiel medzi maximálnou a minimálnou namenanou hodnotou teploty. Jeho veľkosť závisí iba od krajných hodnôt, pričom v tomto prípade spodná hranica patrí medzi odlahlé hodnoty. Preto variačné rozpätie mimo odlahlých hodnôt bude o niečo menšie.

Variačný koeficient je mierou relatívnej variability definovaný ako podiel smerodajnej odchýlky a aritmetického priemeru. Variačný koeficient teploty je 1.01 (101%), čo je dosť veľa a teda dátá sú vysoko variabilné. Ďalej porovnáme variačný koeficient teploty s variačnými koeficientmi iných atribútov, aby sme porovnali ich variabilitu.

Asimetria

Šikmost vyjadruje zošikmenie súboru, teda či prevažujú vysoké alebo nízke hodnoty. Hodnota šiknosti je takmer nulová (-0.00843519) teda hodnoty nie sú zošikmené a ide o symetrické rozdelenie.

Špicatosť vyjadruje rozloženie dát v súbore alebo sústredenie početnosti okolo strednej hodnoty. Hodnota špicatosti je 2.591749, je menšia ako 3 teda hodnoty majú trocha plochšie rozdelenie ako je normálne. To znamená, že v súbore máme viac hodnôt, ktoré sa nachádzajú ďalej od priemeru, resp. veľa nízkych aj vysokých hodnôt, ktoré nie sú blízko priemeru a vrchol rozdelenia početností hodnôt nie je až taký špicatý ako pri normálnom rozdelení.

Boxplot

Graf rozdeľuje namerané hodnoty teploty na niekolko častí. Krabica zobrazuje rozsah v ktorom sa nachádza 50% hodnôt - medzikvartilové rozpätie, čo je hodnota 3.kvartilu - hodnota 1. kvartilu. Hodnota 3.kvartilu je 16.5 a hodnota 1. kvartilu 2, teda medzikvartilové rozpätie (IQR) je $16.5 - 2 = 14.5$. Uprostred krabice je zvýraznený medián hrubou čierrou čiarou. Kedže sa nachádza v strede, dátu nie sú zošikmené ale sú rovnomerne rozptýlené okolo strednej hodnoty.

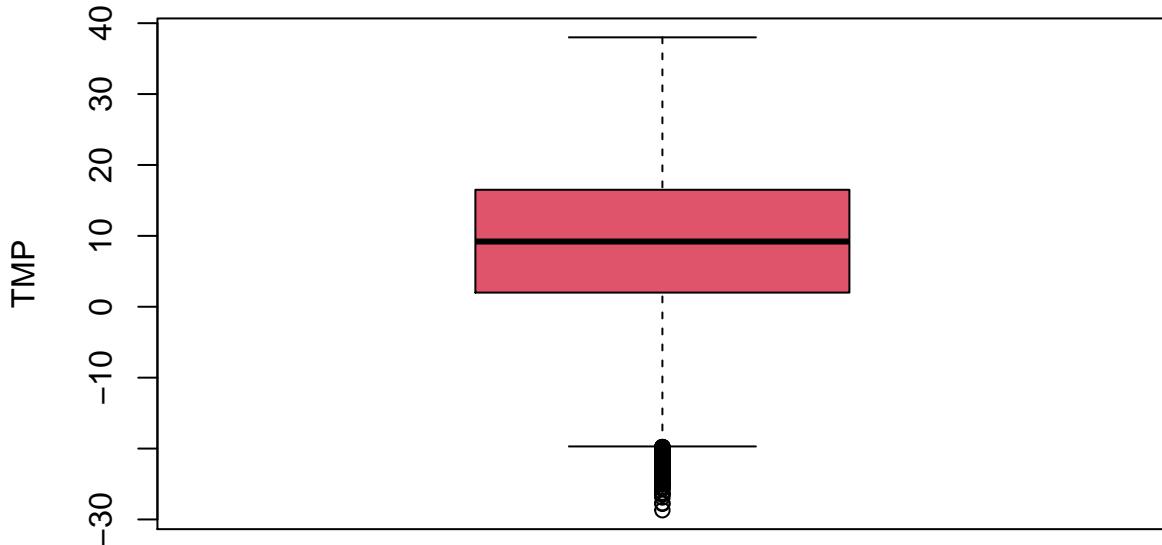
Ďalej z boxplotu vidieť maximálnu a minimálnu hodnotu (vonkajšie hradby boxplotu). Maximálna hodnota (38.25) je vypočítaná ako 3.kvartil + $1.5 * \text{IQR}$ (medzikvartilové rozpätie). Minimálna hodnota (-19.75) je vypočítaná ako 1.kvartil - $1.5 * \text{IQR}$ (medzikvartilové rozpätie).

Všetky hodnoty nachádzajúce sa nad a pod maximálnou a minimálnou hodnotou môžme považovať za odlahlé hodnoty. Počet odlahlých hodnôt pod minimálnou hodnotou je 403, nad maximálnou sa nenachádzajú žiadne odlahlé hodnoty.

Nakoniec pre odlahlé hodnoty overíme či patria medzi extrémne. Horná hranica extrémnych hodnôt je vypočítaná ako 3.kvartil + $3 * \text{IQR}$. Dolná hranica extrémnych hodnôt je vypočítaná ako 1.kvartil - $3 * \text{IQR}$. Takéto hodnoty sa v našich dátach nevyskytujú.

```
boxplot(all_data$TMP, col = 2, ylab = "TMP", main = "Temperature boxplot")
```

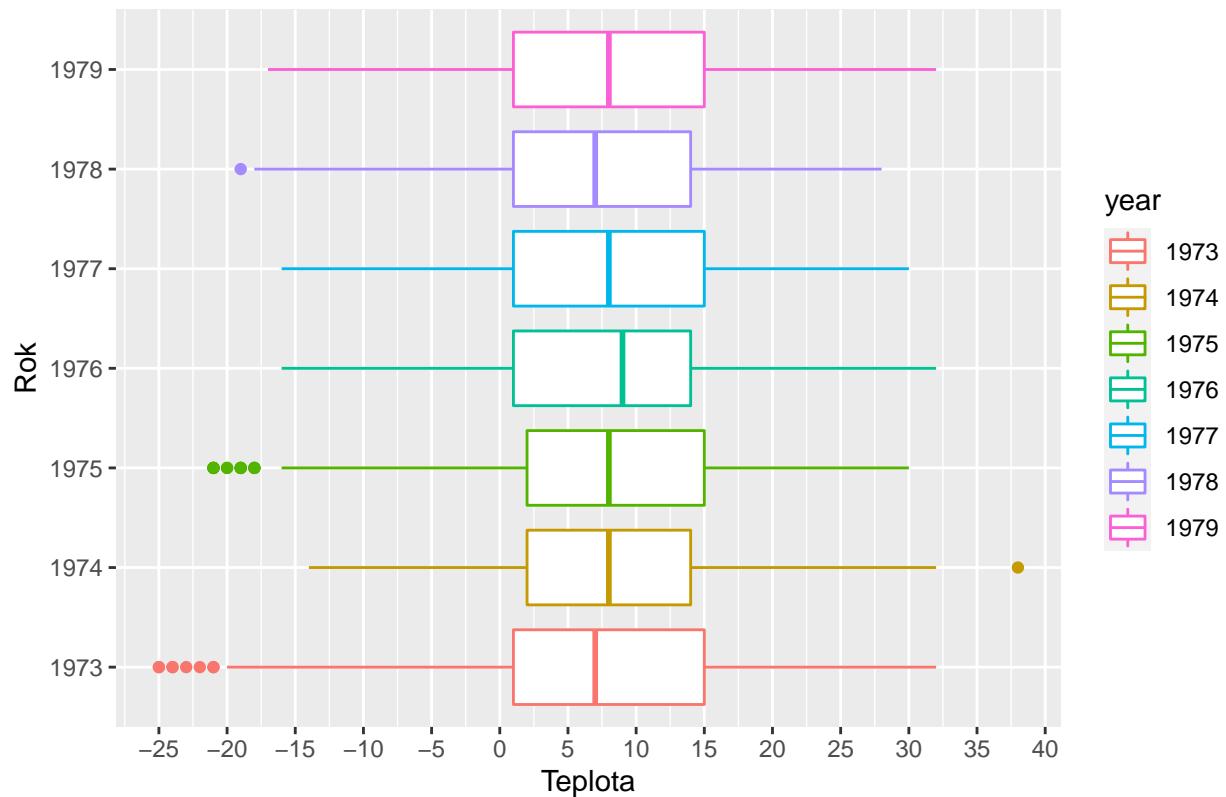
Temperature boxplot



```
df <- all_data %>%
  dplyr::mutate(
    year = ymd_hms(DATE) %>%
      lubridate::year() %>%
      map_chr(~ as.character(.x))
  ) %>%
  dplyr::select(all_of(c('year', 'TMP')))

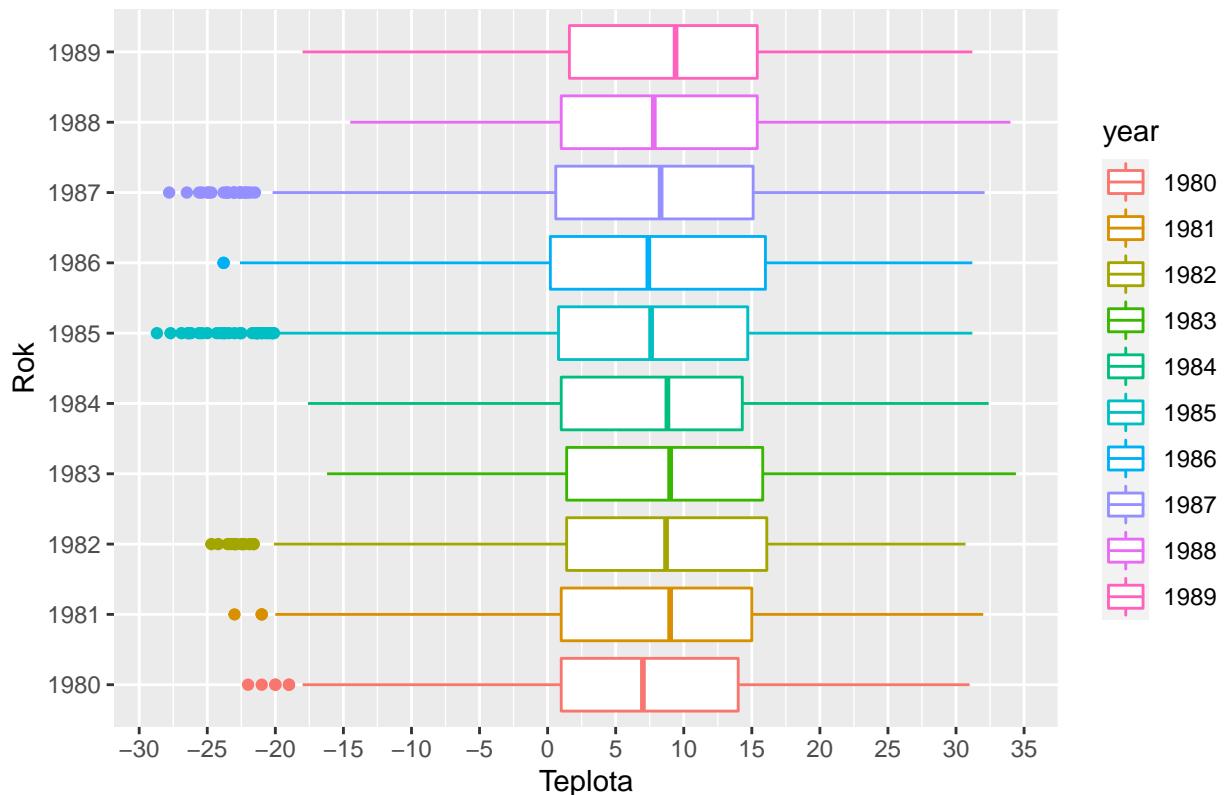
ggplot(data = df %>% filter(year < 1980), aes(TMP, factor(year), colour = year)) +
  geom_boxplot() +
  labs(title = paste("Boxploty teploty pre roky 1973 - 1979"), x = "Teplota", y = "Rok", fill = "year")
  scale_x_continuous(breaks = seq(-40, 40, by = 5))
```

Boxploty teploty pre roky 1973 – 1979



```
ggplot(data = df %>% filter(year >= 1980 & year < 1990), aes(TMP, factor(year), colour = year)) +
  geom_boxplot() +
  labs(title = paste("Boxploty teploty pre roky 1980 – 1989"), x = "Teplota", y = "Rok", fill = "year") +
  scale_x_continuous(breaks = seq(-40, 40, by = 5))
```

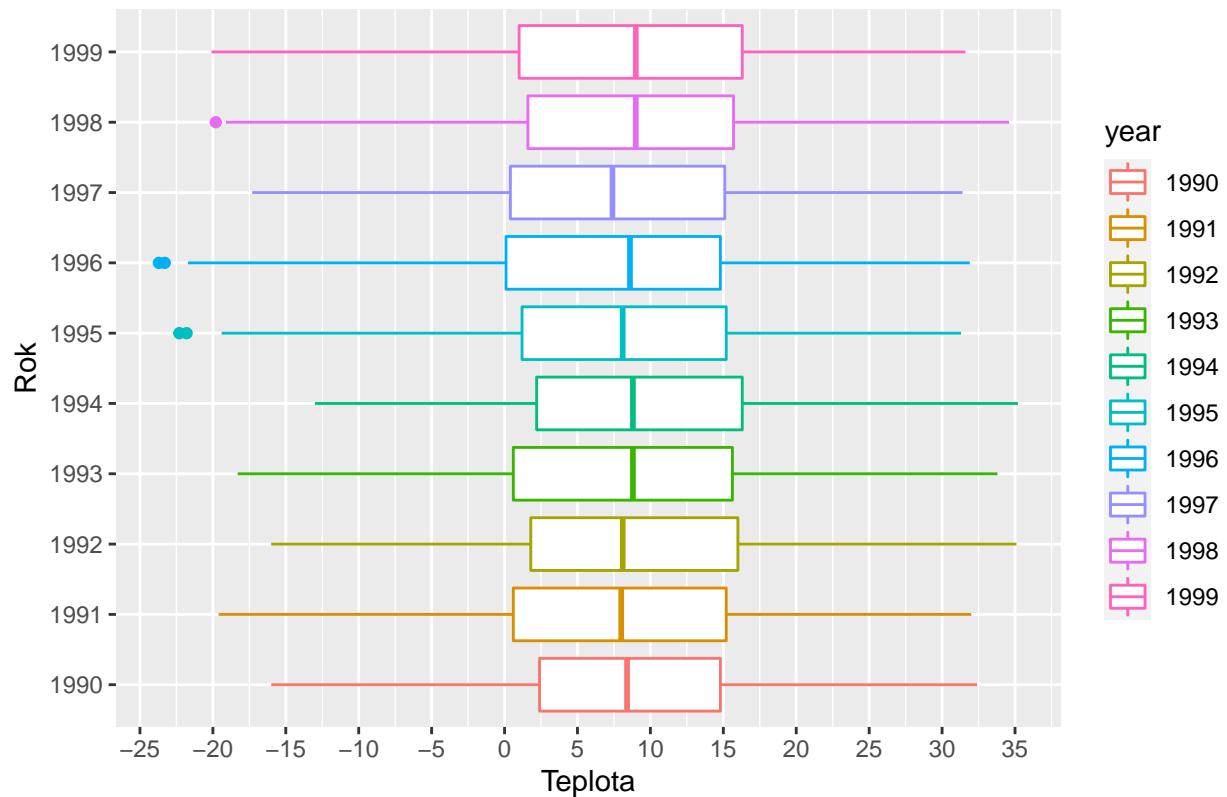
Boxploty teploty pre roky 1980 – 1989



Pri porovnaní 70-tych a 80-tych rokov vidno mierne zvýšenie teploty, medián pre 80-te roky je posunutý bližie k 10°C, pričom v 70-tych rokoch je približne okolo 7.5°C.

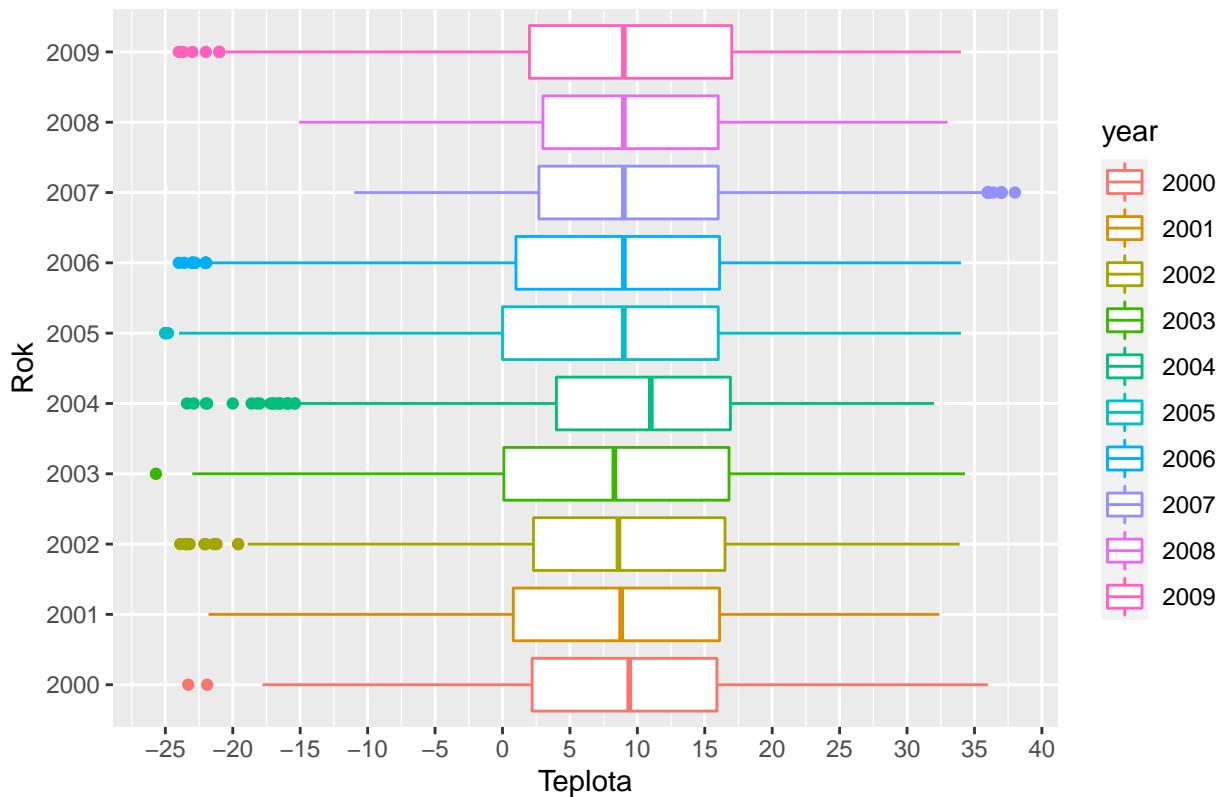
```
ggplot(data = df %>% filter(year >= 1990 & year < 2000), aes(TMP, factor(year), colour = year)) +
  geom_boxplot() +
  labs(title = paste("Boxploty teploty pre roky 1990 – 1999"), x = "Teplota", y = "Rok", fill = "year")
  scale_x_continuous(breaks = seq(-40, 40, by = 5))
```

Boxploty teploty pre roky 1990 – 1999



```
ggplot(data = df %>% filter(year >= 2000 & year < 2010), aes(TMP, factor(year), colour = year)) +
  geom_boxplot() +
  labs(title = paste("Boxploty teploty pre roky 2000 - 2009"), x = "Teplota", y = "Rok", fill = "year")
  scale_x_continuous(breaks = seq(-40, 40, by = 5))
```

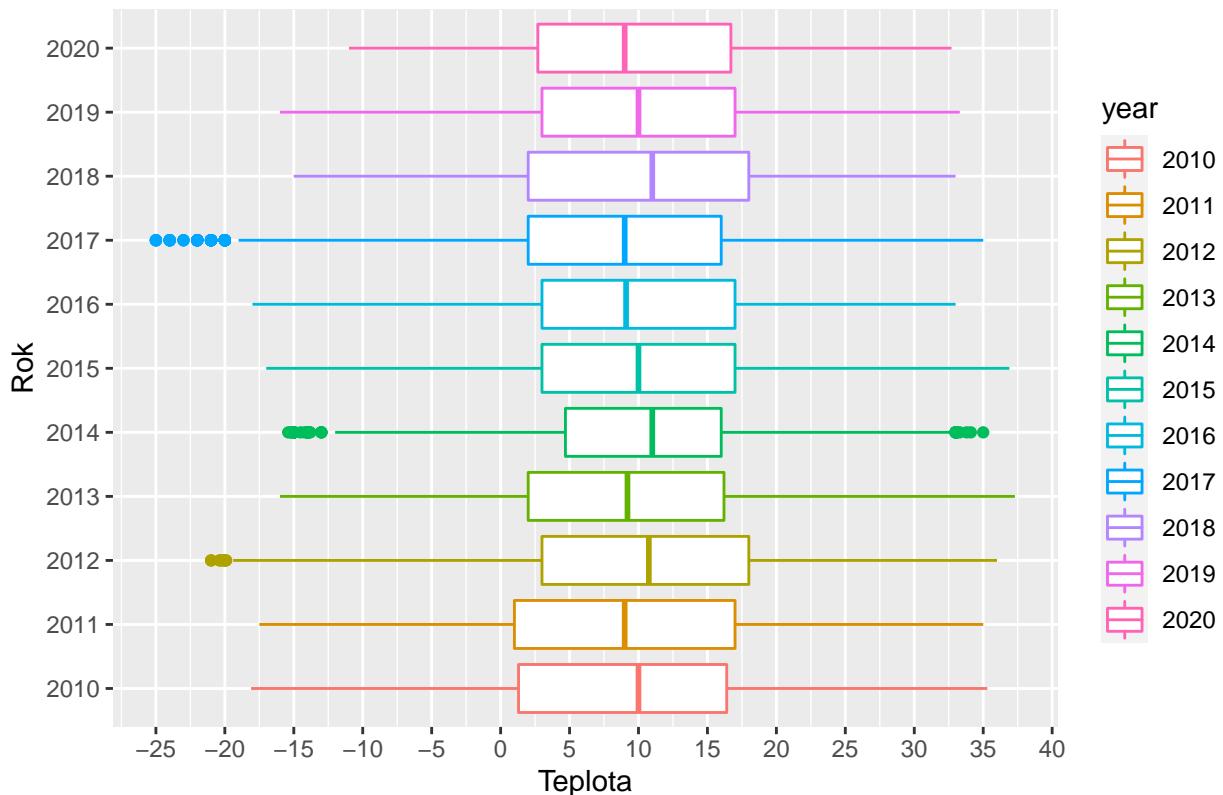
Boxploty teploty pre roky 2000 – 2009



Z nasledujúcich boxplotov pre jednotlivé roky vidíme, že v roku 2017 boli namerané aj veľmi nízke hodnoty teploty. Naopak v roku 2014 má krabica boxplotu rozptyl najmenší, teda v tomto roku boli hodnoty dosť konzistentné oproti iným rokom. V porovnaní so 70-tými a 80-tými rokmi sa krabica boxplotov posunula približne o 2-3°C k vyšším teplotám, z čoho by sme tiež mohli povedať, že s časom sa teplota zvyšuje.

```
ggplot(data = df %>% filter(year >= 2010), aes(TMP, factor(year), colour = year)) +
  geom_boxplot() +
  labs(title = paste("Boxploty teploty pre roky 2010 – 2020"), x = "Teplota", y = "Rok", fill = "year")
  scale_x_continuous(breaks = seq(-40, 40, by = 5))
```

Boxploty teploty pre roky 2010 – 2020



Grafy priemerných mesačných teplôt pre rok 2010 a 2020. Vidíme mierne zvýšenie teploty koncom leta a počas jesenných mesiacov. Taktiež priemerná teplota v decembri je v roku 2020 nad 0.

```
all_data_split_date <- mutate(all_data, day = mday(DATE), month = month(DATE), year = year(DATE), hour = hour(DATE))

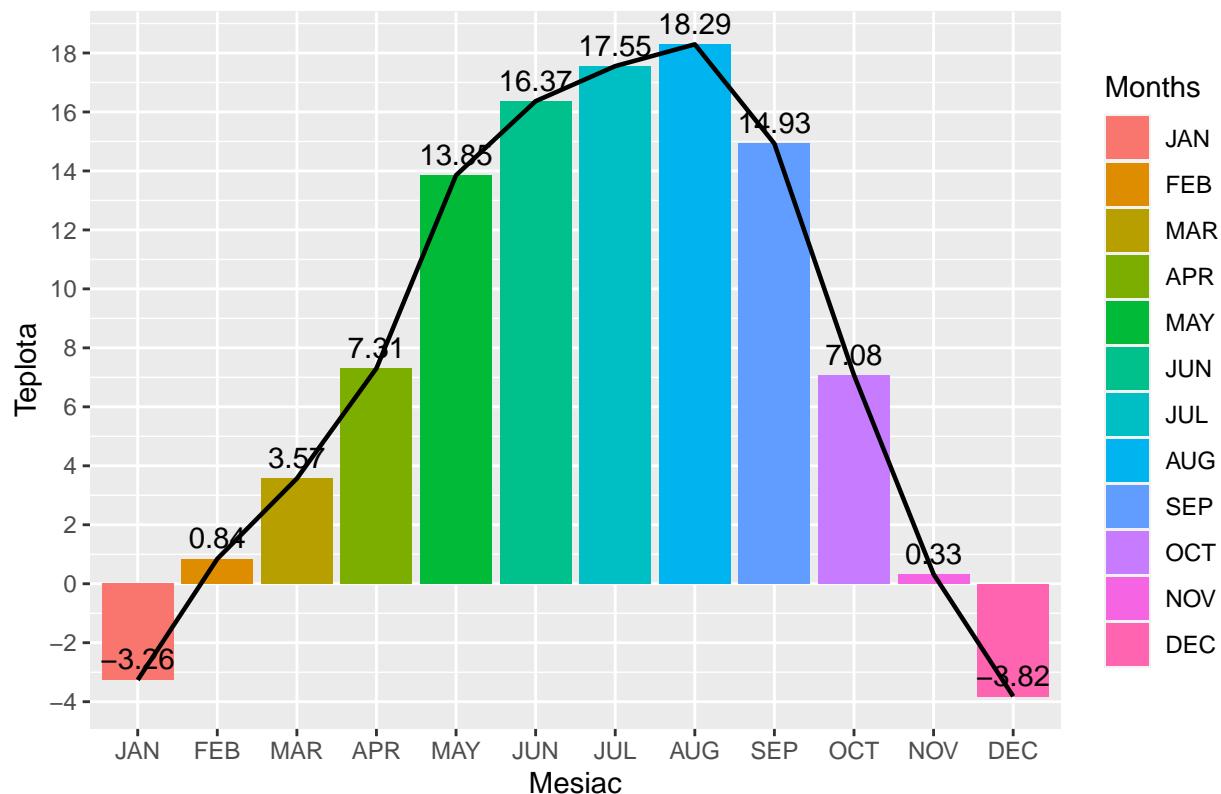
all_data_split_date <- all_data_split_date[!is.na(all_data_split_date$TMP), ]

months <- c('JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC')

# Mean TMP 1973
mean_tmp_1973 <- filter(all_data_split_date, year == 1973) %>% group_by(month) %>% summarise(mean = mean(TMP))

ggplot(mean_tmp_1973, aes(x = factor(months, months), y = mean)) +
  geom_col(mapping = aes(fill = factor(months, months))) +
  geom_text(mapping = aes(label = round(mean, 2)), vjust = -0.5) +
  geom_line(mapping = aes(x = month, y = mean), size = 0.8) +
  labs(title = paste("Priemerná mesačná teplota v roku 1973"), x = "Mesiac", y = "Teplota",
       fill = "Months") +
  scale_y_continuous(breaks = seq(-10, 30, by = 2))
```

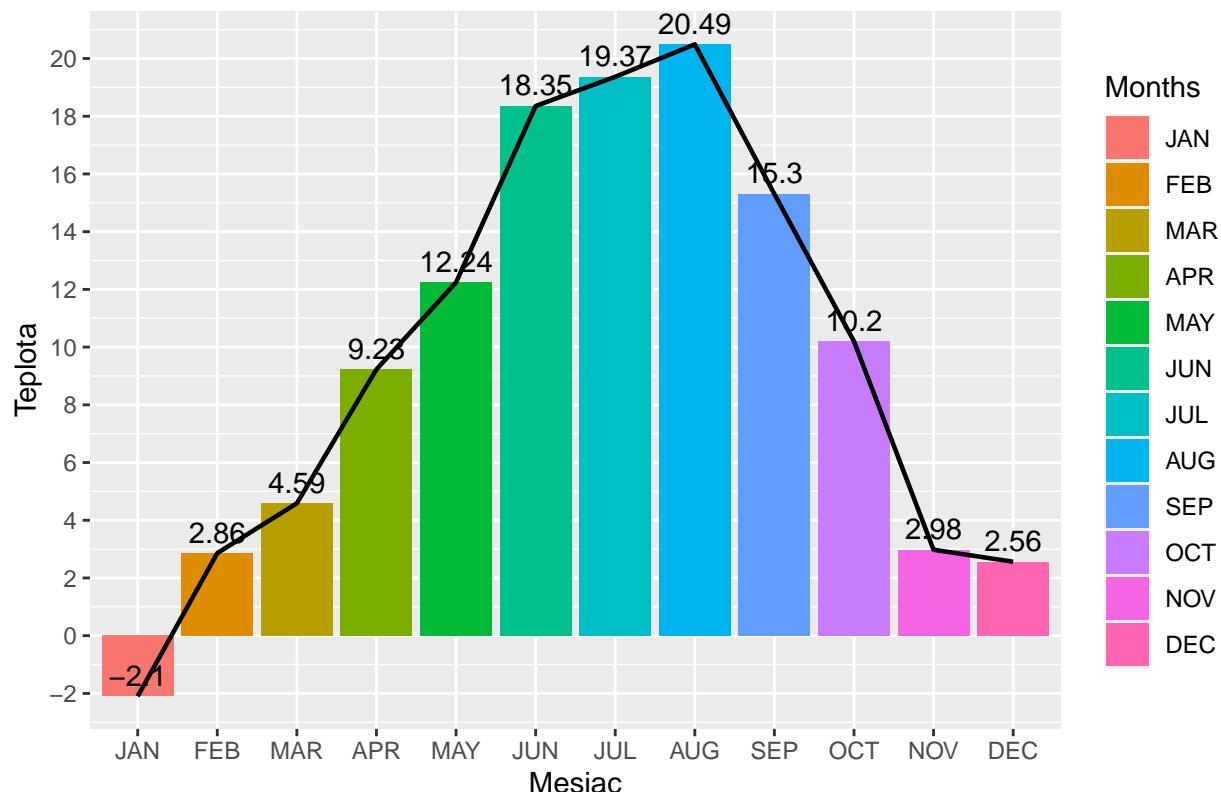
Priemerná mesacná teplota v roku 1973



```
# Mean TMP 2020
mean_tmp_2020 <- filter(all_data_split_date, year == 2020) %>% group_by(month) %>% summarise(mean = mean(temp))

ggplot(mean_tmp_2020, aes(x = factor(months,months), y = mean)) +
  geom_col(mapping = aes(fill = factor(months,months))) +
  geom_text(mapping = aes(label = round(mean,2)), vjust = -0.5) +
  geom_line(mapping = aes(x = month, y = mean), size = 0.8) +
  labs(title = paste("Priemerná mesačná teplota v roku 2020"), x = "Mesiac", y = "Teplota",
       fill = "Months") +
  scale_y_continuous(breaks = seq(-10, 30, by = 2))
```

Priemerná mesacná teplota v roku 2020

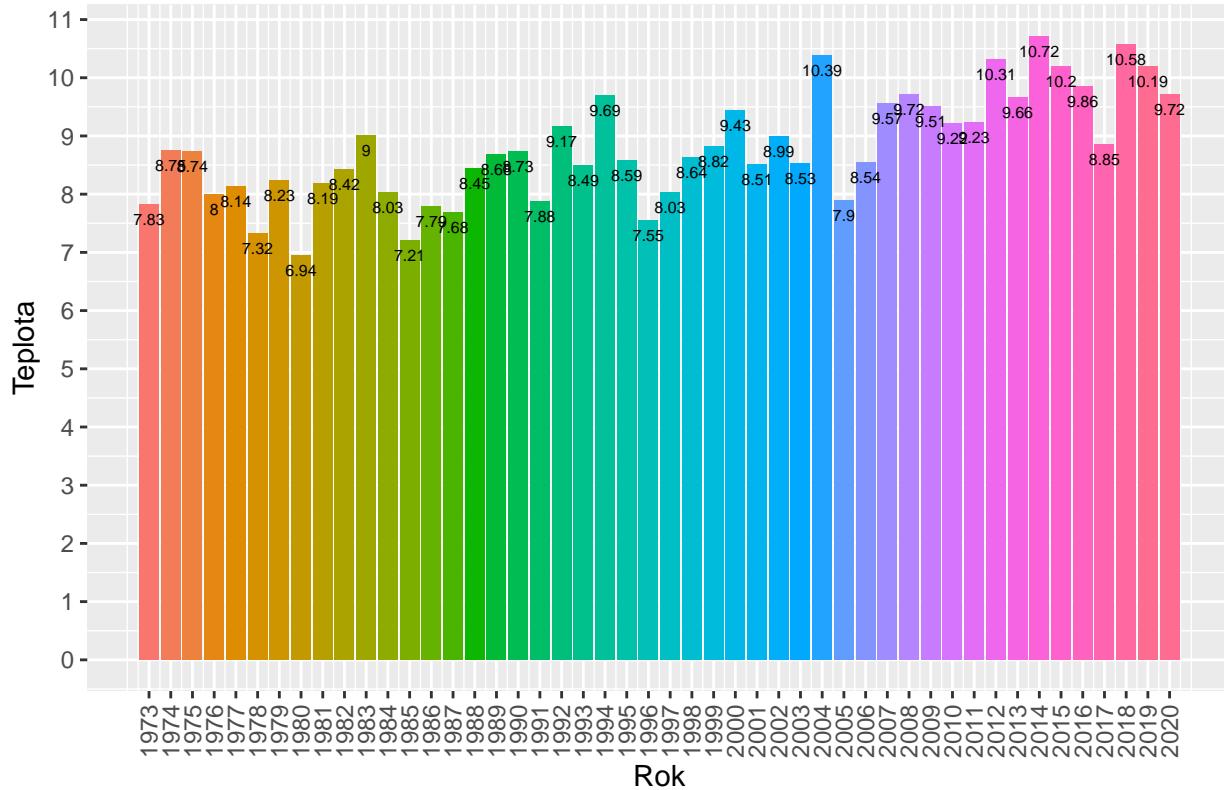


Porovnanie priemerných mesačných teplôt v roku 1973 a 2020. Z grafov vidieť, že sa zvýšila priemerná teplota v zimných mesiacoch nad 0°C (okrem januára). Celkovo sa priemerné teploty zvýšili približne o 1-2°C.

```
# Mean TMP by year
mean_yr_tmp <- all_data_split_date %>% group_by(year) %>% summarise(mean = mean(na.omit(TMP)))

ggplot(mean_yr_tmp, aes(x = year, y = mean)) +
  geom_col(mapping = aes(fill = factor(year,year))) +
  geom_text(mapping = aes(label = round(mean,2)), vjust = 1.8, size = 2) +
  labs(title = paste("Priemerná ročná teplota pre roky 1973 - 2020"), x = "Rok", y = "Teplota",
       fill = "Years") +
  scale_y_continuous(breaks = seq(0, 20, by = 1)) +
  scale_x_continuous(breaks = seq(1973, 2020, by = 1)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  theme(legend.position="none")
```

Priemerná ročná teplota pre roky 1973 – 2020

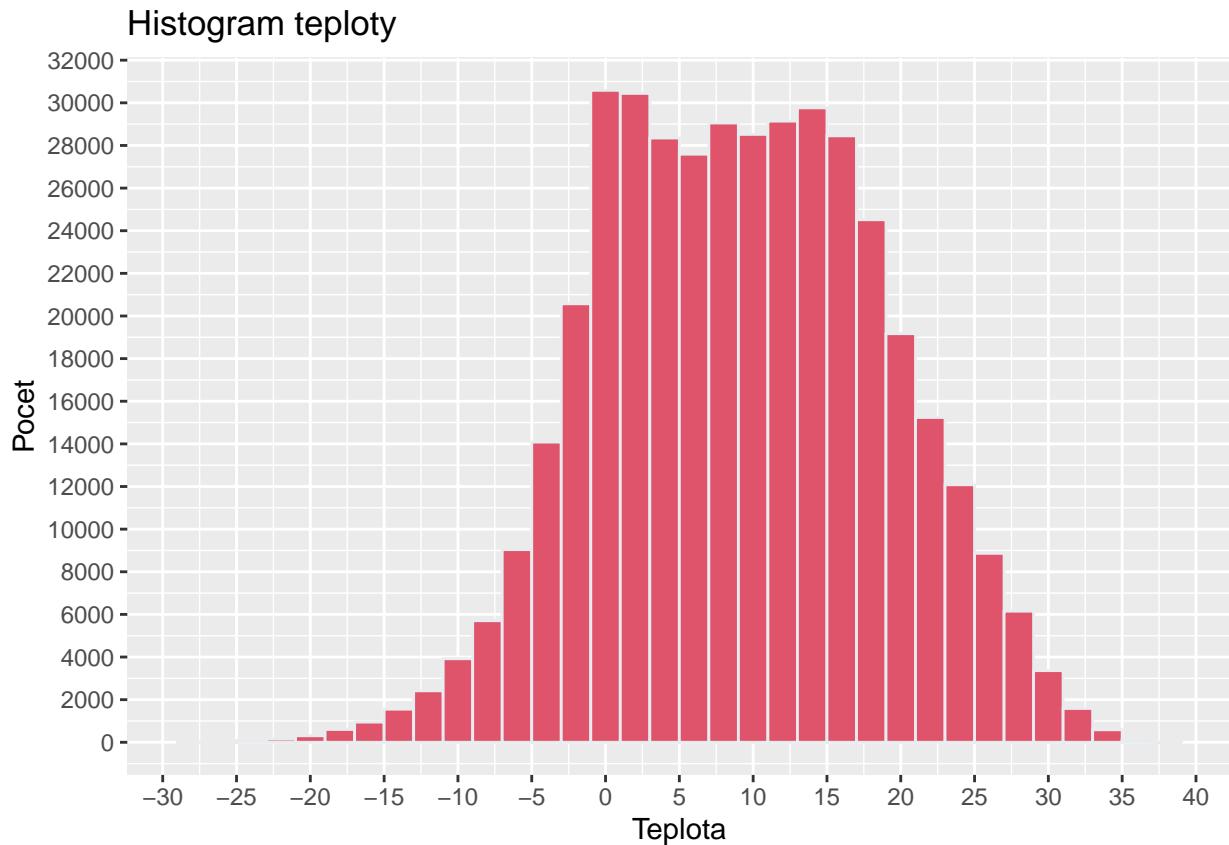


Graf priemerných ročných teplôt pre roky 1973 - 2020. Vidíme, že najnižšia priemerná teplota bola v roku 1980 a najvyššia v roku 2014. Celkovo môžeme vidieť mierne zvýšenie priemerných teplôt v rokoch 2010 - 2020 v porovnaní so 70.-tymi rokmi.

Histogram

Na histograme vidíme početnosti jednotlivých hodnôt teploty. Najpočetnejšie sú hodnoty medzi 0 - 20°C. Rozdelenie dát sa podobá normálnemu rozdeleniu. Hodnoty sú symetrické okolo mediánu.

```
ggplot(all_data, aes(x = TMP)) +
  geom_histogram(bins = 40, binwidth = 2, fill = 2, color = "#e9ecf") +
  labs(title = paste("Histogram teploty")) +
  xlab("Teplota") +
  ylab("Počet") +
  scale_x_continuous(breaks = seq(-40, 40, by = 5)) +
  scale_y_continuous(breaks = seq(0, 40000, by = 2000))
```



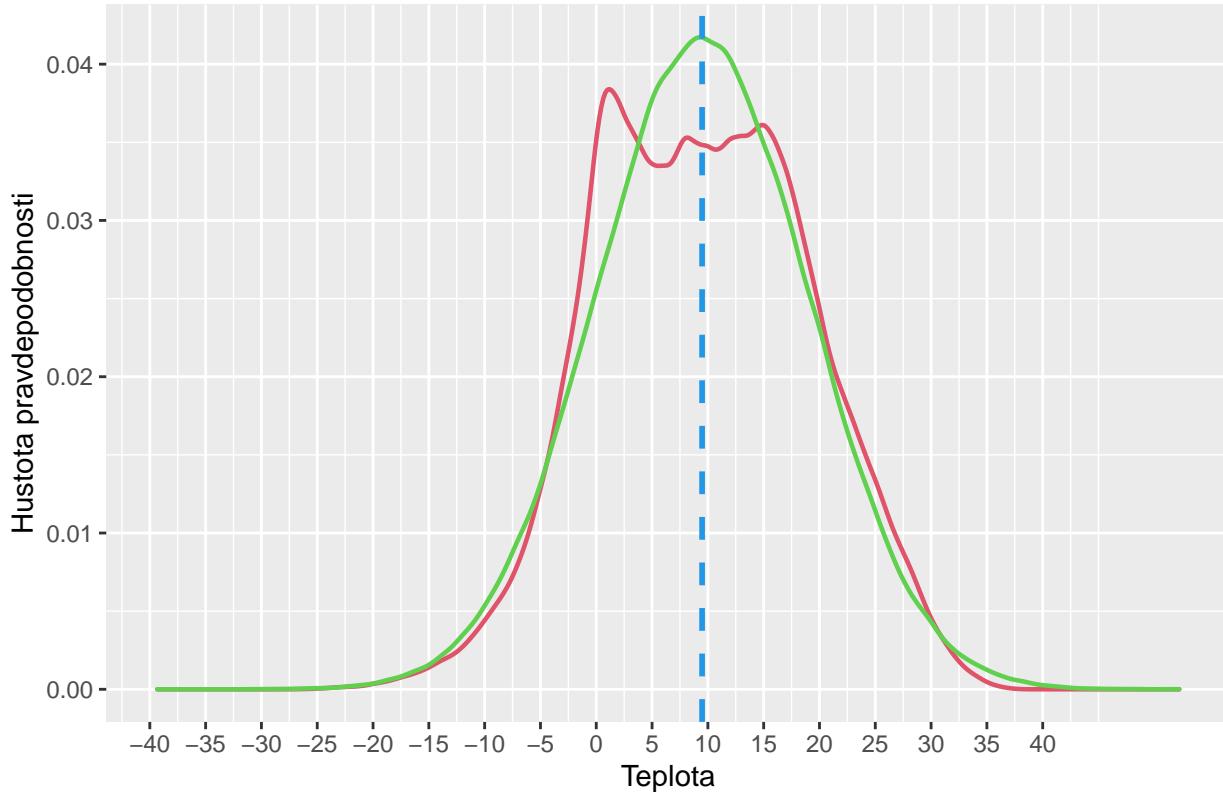
Graf hustoty

Graf hustoty slúži na porovnanie priebehu hustoty pravdepodobnosti normálneho rozdelenia (zelená čiara) a odhadu hustoty vypočítaného z namerných hodnôt teploty (červená čiara). Čiary nie sú rovnaké, vrch rozdelenia je mierne zvlnený oproti normálnemu rozdeleniu. Modrá reprešovaná čiara predstavuje priemernú teplotu. Krivka teploty je viac plochá ako krivka normálneho rozdelenia, o čom svedčí aj hodnota špicatosti, ktorá je o niečo nižšia ako hodnota špicatosti normálneho rozdelenia. Z grafu vidno aj vrchol - najčastejšiu hodnotu 1 a niekoľko menších vrcholov v okolí hodnoty 8 a 15.

```
# denisty plot
# data z normalneho rozdelenia
data_norm <- data.frame(dens = c(rnorm(length(na.omit(all_data$TMP))), mean(all_data$TMP, na.rm = T), sd

# porovnanie hodnot normalneho rozdelenia a TMP
ggplot(all_data, aes(x = TMP), color = 3) +
  geom_density(color = 2, size = 0.8) +
  geom_density(data_norm, mapping = aes(x = dens), color = 3, size = 0.8) +
  geom_vline(aes(xintercept = mean(TMP, na.rm = T)),
             color = 4, linetype = "dashed", size = 1) +
  scale_x_continuous(breaks = seq(-40, 40, by = 5)) +
  labs(title = paste("Odhad hustoty teploty")) +
  xlab("Teplota") +
  ylab("Hustota pravdepodobnosti")
```

Odhad hustoty teploty



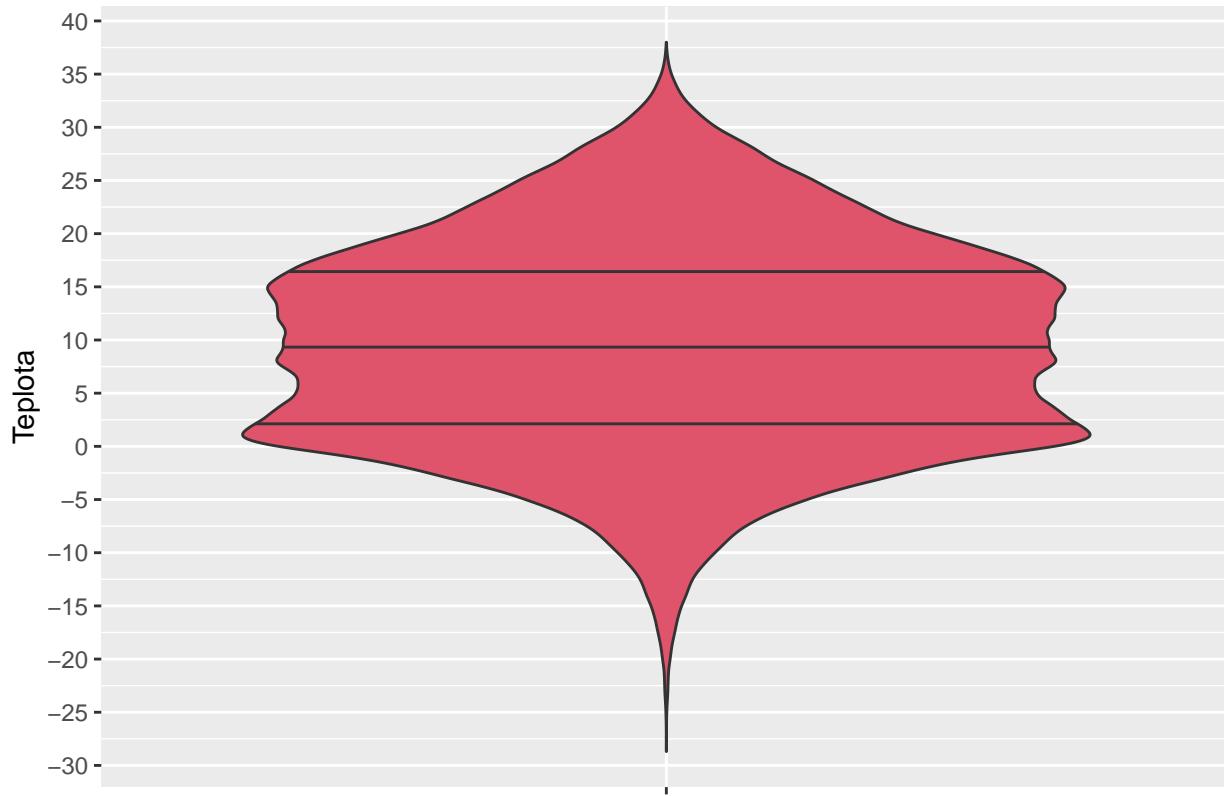
Husľový graf

Husľový graf doplnený o hlavné kvartily zobrazuje rozdelenie hustoty, pričom aj podľa tohto grafu vidíme, že ide o normálne rozdelenie. Dáta sú najpočetnejšie v strede približne okolo hodnoty 1. a 3. kvartílu a smerom k vyšším, resp. nižším hodnotám sa ich hustota zmenšuje.

```
df <- all_data %>%
  dplyr::select('TMP') %>%
  tidyr::gather(key = 'label', value = 'tmp')

ggplot(data = df, aes(factor(label), tmp, fill = tmp)) +
  geom_violin(draw_quantiles = c(0.25, 0.5, 0.75), fill = 2) +
  labs(title = paste("Husľový graf teploty"), y = "Teplota", fill = "temperature") +
  theme(axis.title.x = element_blank()) +
  theme(axis.text.x = element_blank()) +
  scale_y_continuous(breaks = seq(-40, 40, by = 5))
```

Huslový graf teploty

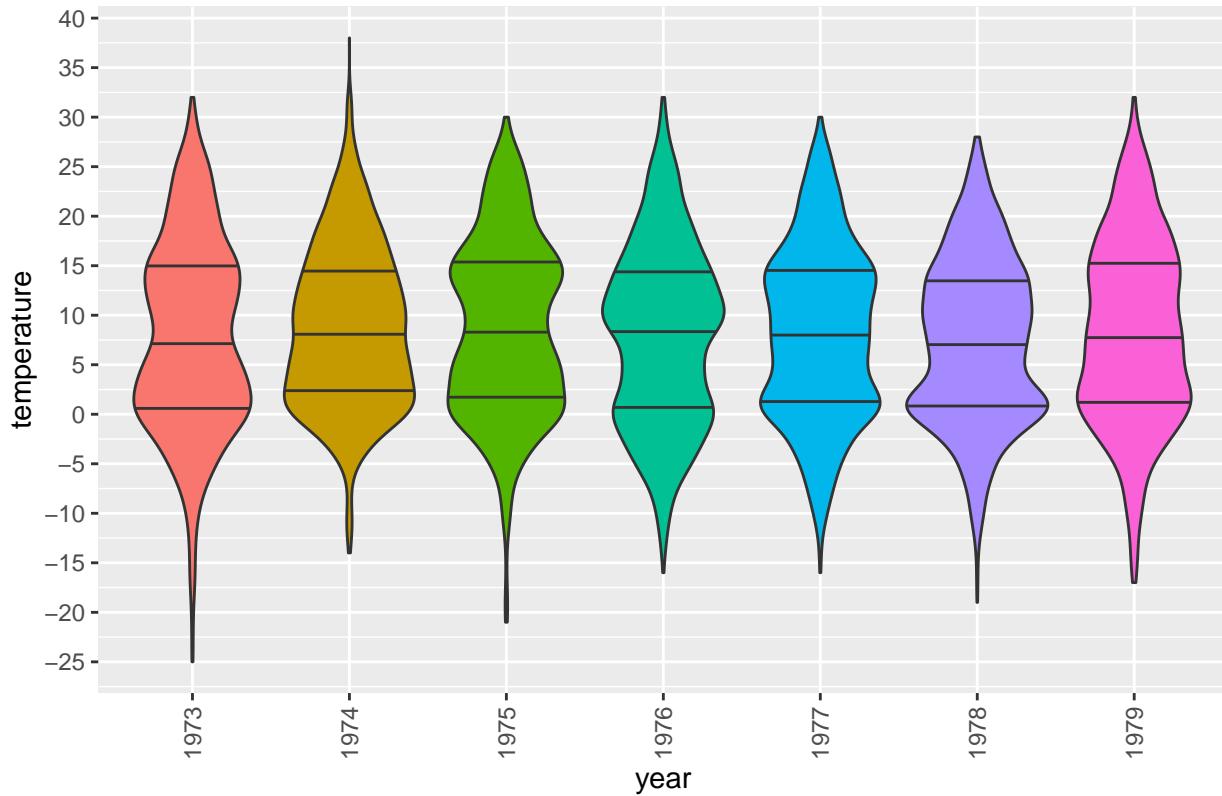


Z nasledujúcich grafov vidíme rozdelenie hustoty pre jednotlivé roky 1973 - 1979 a 2010 - 2020. V roku 2017 boli dátá najmenej konzistentné a obsahujú viac nízkych hodnôt ako po iné roky. Hodnoty kvartilov sú v porovnaní zo 70-tymi rokmi posunuté vyššie.

```
df <- all_data %>%
  dplyr::mutate(
    year = ymd_hms(DATE) %>%
      lubridate::year() %>%
      map_chr(~ as.character(.x))
  ) %>%
  dplyr::select(all_of(c('year', 'TMP')))

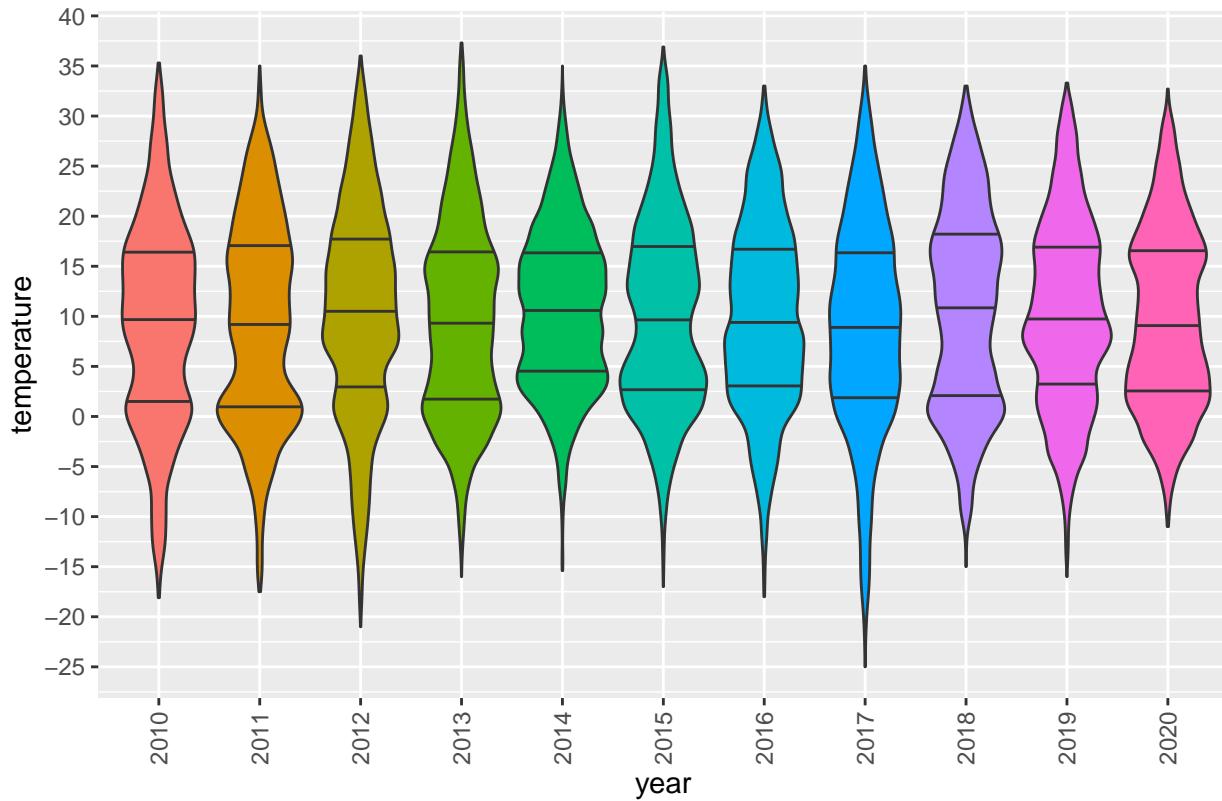
ggplot(data = df %>% filter(year < 1980), aes(factor(year), TMP, fill = year)) +
  geom_violin(draw_quantiles=c(0.25, 0.5, 0.75)) +
  labs(title = paste("Huslový graf teploty pre roky 1973 - 1979"), x = "year", y = "temperature", fill =
  scale_y_continuous(breaks = seq(-40, 40, by = 5)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  theme(legend.position="none")
```

Huslový graf teploty pre roky 1973 – 1979



```
ggplot(data = df %>% filter(year >= 2010), aes(factor(year), TMP, fill = year)) +  
  geom_violin(draw_quantiles=c(0.25, 0.5, 0.75)) +  
  labs(title = paste("Huslový graf teploty pre roky 2010 – 2020"), x = "year", y = "temperature", fill =  
    scale_y_continuous(breaks = seq(-40, 40, by = 5)) +  
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +  
    theme(legend.position="none")
```

Huslový graf teploty pre roky 2010 – 2020



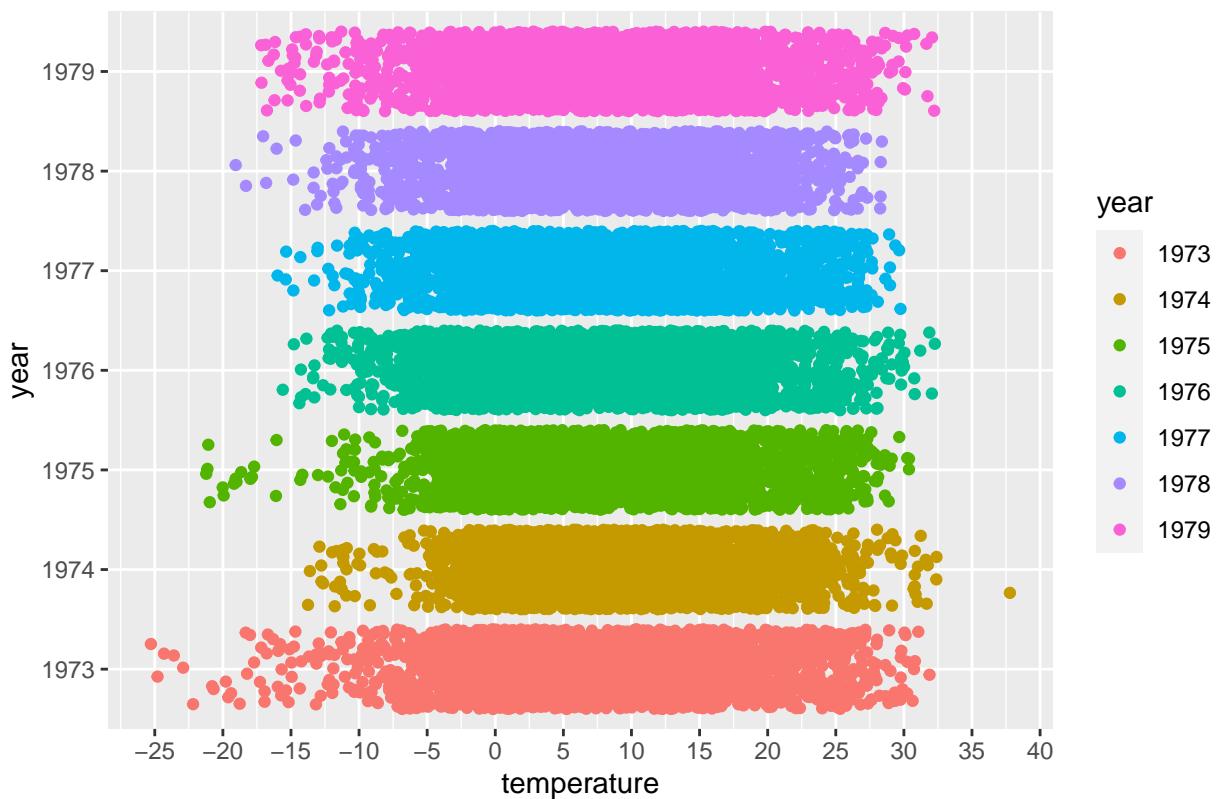
Graf rozptylenia

Na osi x sú vynesené namerané hodnoty teploty a os y reprezentuje jednotlivé roky. Vidíme, že rozloženie dát je najhustejsie medzi hodnotami 0 - 20. Čím nižšia je hodnota teploty tým sa aj počet nameraných hodnôt znižuje, podobne aj pri vyokých teplotách. Ďalej vidíme, že v roku 2017 bola pravdepodobne veľmi chladná zima, pretože v tomto roku teplota dosahuje najnižšie hodnoty (mínus 20°C až mínus 25°C). Naopak v roku 2020 je len veľmi málo nízkych hodnôt (maximálne po mínus 10°C). V roku 2017 sú merané iba celočíselné hodnoty teploty. V porovnaní so 70-tymi orkmi je väčšia hustota dát aj pri teplotách nad 20°C.

```
df <- all_data %>%
  dplyr::mutate(
    year = ymd_hms(DATE) %>%
      lubridate::year() %>%
      map_chr(~ as.character(.x))
  ) %>%
  dplyr::select(all_of(c('year', 'TMP')))

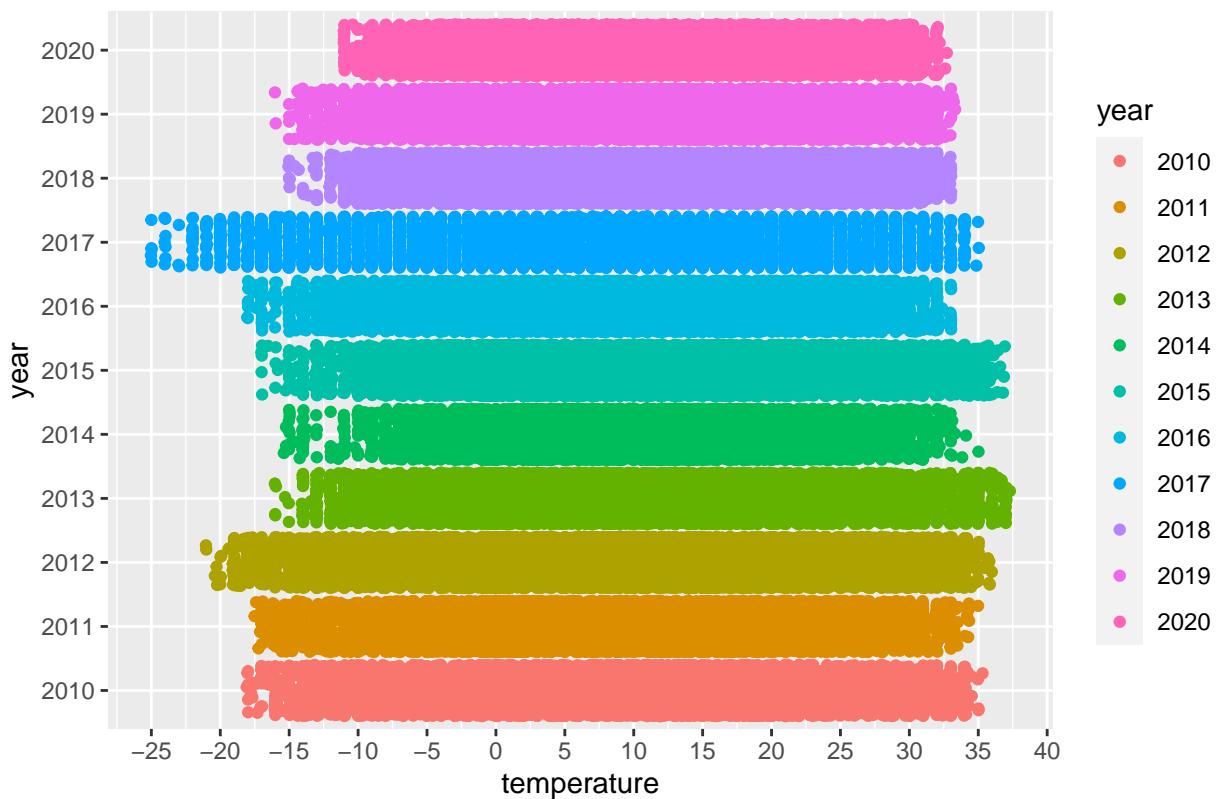
ggplot(data = df %>% filter(year < 1980), aes( TMP,factor(year), colour=year)) +
  geom_jitter() +
  labs(title = paste("Graf rozptylenia teploty pre roky 1973 – 1979"), x = "temperature", y = "year", f
```

Graf rozptylenia teploty pre roky 1973 – 1979



```
ggplot(data = df %>% filter(year >= 2010), aes( TMP,factor(year), colour=year)) +  
  geom_jitter() +  
  labs(title = paste("Graf rozptylenia teploty pre roky 2010 – 2020"), x = "temperature", y = "year", f
```

Graf rozptýlenia teploty pre roky 2010 – 2020

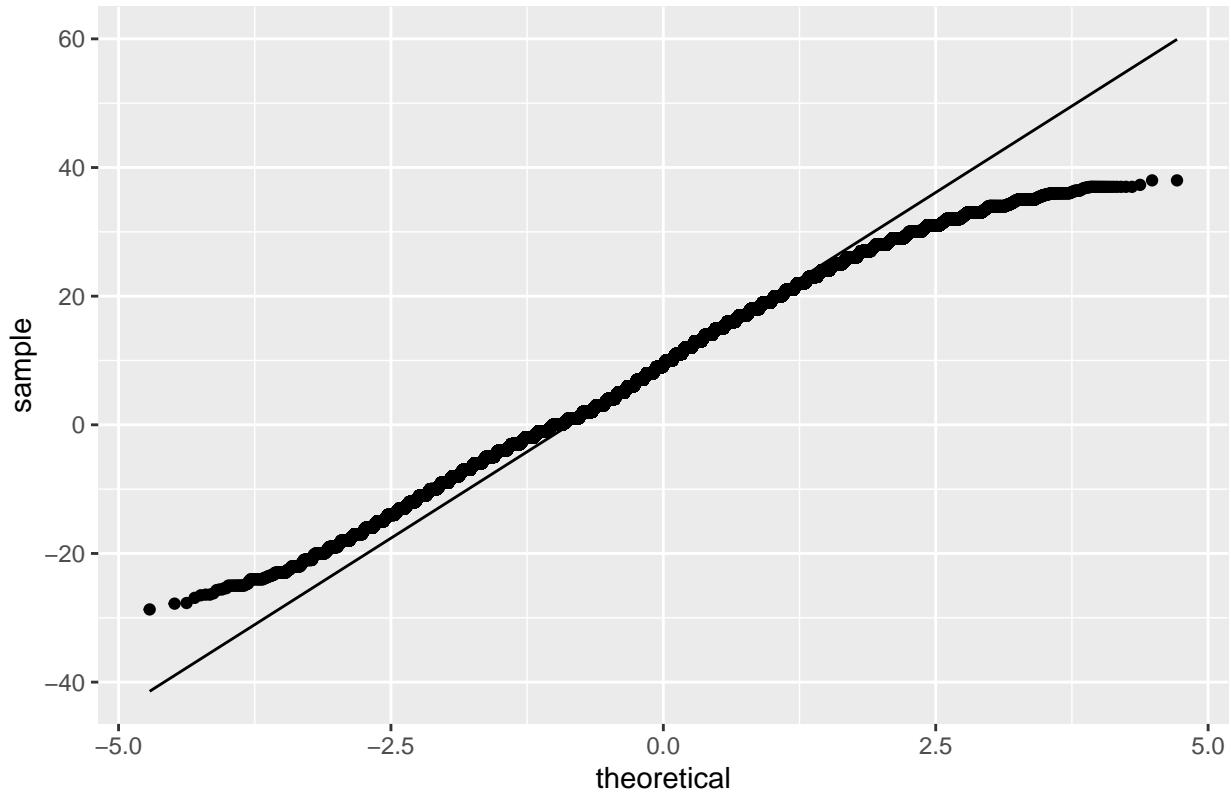


Q-Q graf

Graf zobrazuje odchýlku empirického od teoretického normálneho rozdelenia. Empirické rozdelenie je rozdelenie nameraných hodnôt teploty. Body ležia veľmi blízko priamky, odchylujú sa len body v oblasti horných a dolných kvantilov, teda rozdelenie hodnôt teploty je podobné normálnemu rozdeleniu.

```
ggplot(data = all_data, aes(sample = TMP)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = paste("Q-Q graf teploty"))
```

Q–Q graf teploty



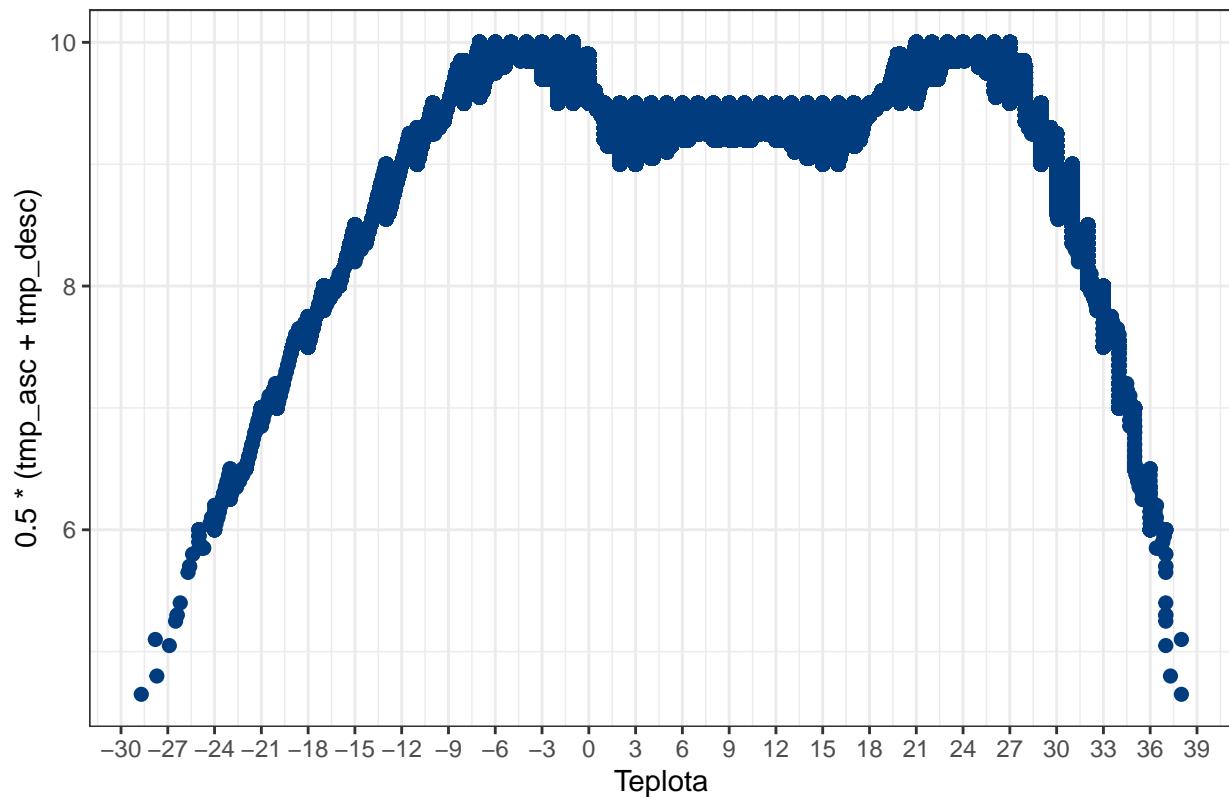
Graf polosum

Z grafu polosum vidno, že hodnoty sú takmer symetrické okolo mediánu. Mierne sa odchyľujú len krajné hodnoty (veľmi nízke a veľmi vysoké).

```
# polosum
tmp <- all_data$TMP
tmp_asc <- sort(tmp, decreasing = FALSE)
tmp_desc <- sort(tmp, decreasing = TRUE)

ggplot(data.frame(tmp_asc), aes(x = tmp_asc, y = 0.5*(tmp_asc+tmp_desc))) +
  geom_point(size = 2, color = "#013c7f") +
  scale_x_continuous(breaks = seq(-30, 40, by = 3)) +
  labs(title = "Graf polosum pre teplotu", x = "Teplota") +
  theme_bw()
```

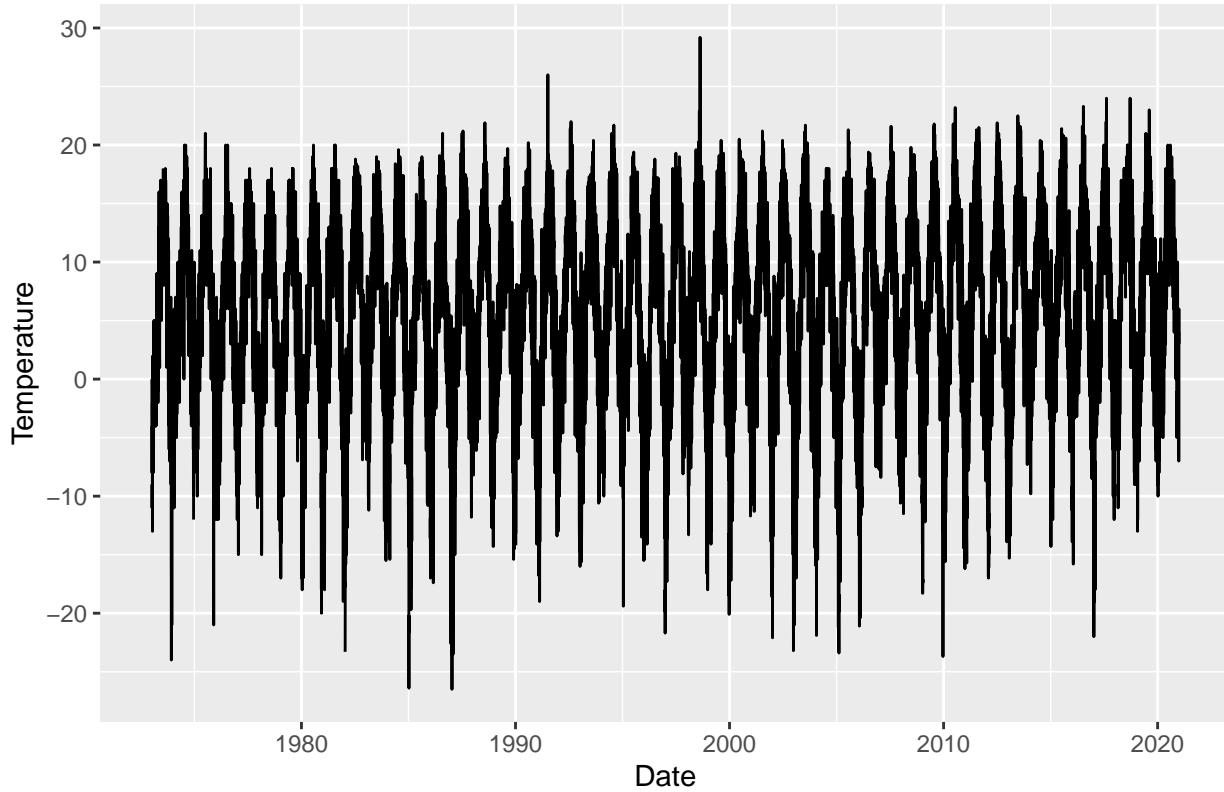
Graf polosum pre teplotu



Časový graf teploty

```
all_data %>%
  dplyr::mutate(
    date = as_date(DATE)
  ) %>%
  dplyr::distinct(date, .keep_all=TRUE) %>%
  dplyr::select(date, TMP) %>%
  as_tsibble(
    index = date
  ) %>%
  autoplot( TMP) +
  labs(title = "Time graph of temperature",
       y = "Temperature", x = "Date")
```

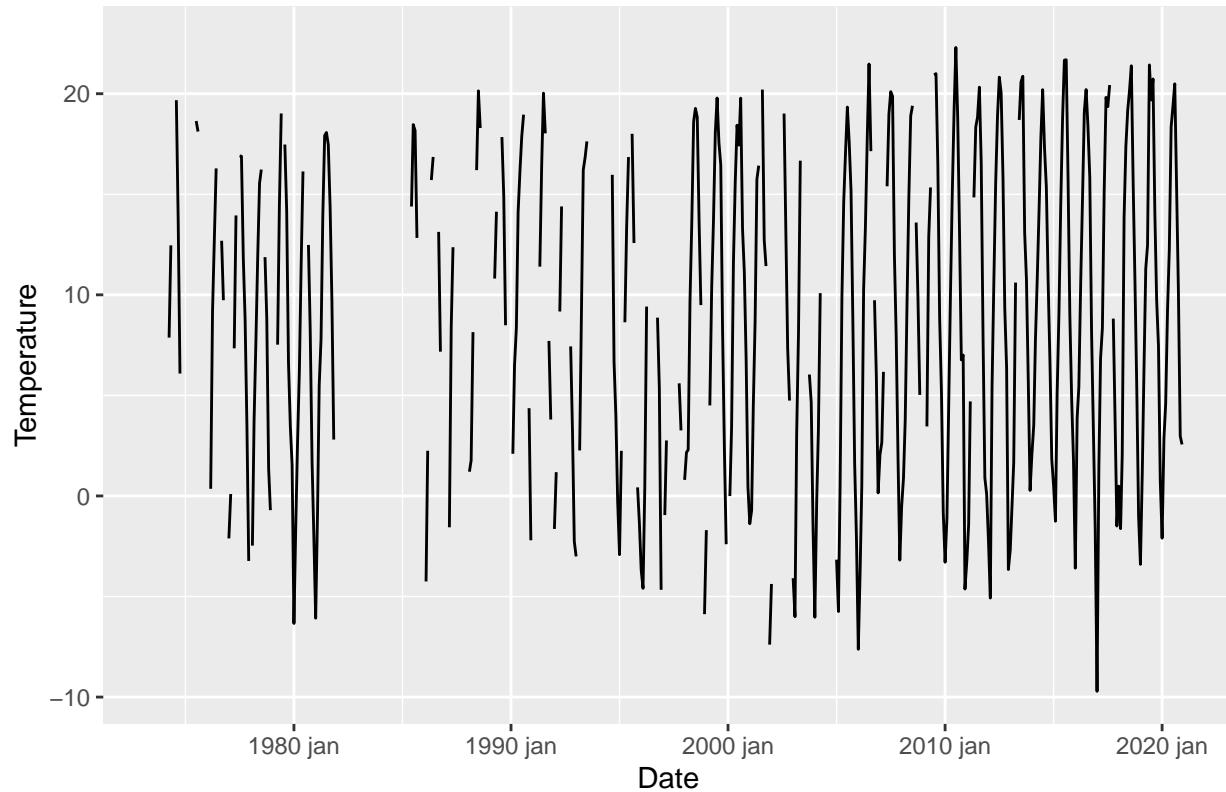
Time graph of temperature



```
all_data %>%
  dplyr::mutate(
    year_month = yearmonth(DATE)
  ) %>%
  dplyr::group_by(year_month) %>%
  dplyr::summarise(tmp = na.omit(mean(TMP))) %>%
  as_tsibble(
    index = year_month
  ) %>%
  tsibble::fill_gaps() %>%
  autoplot( tmp) +
  labs(title = "Time graph of temperature",
       y = "Temperature", x = "Date")

## `summarise()` has grouped output by 'year_month'. You can override using the `groups` argument.
```

Time graph of temperature



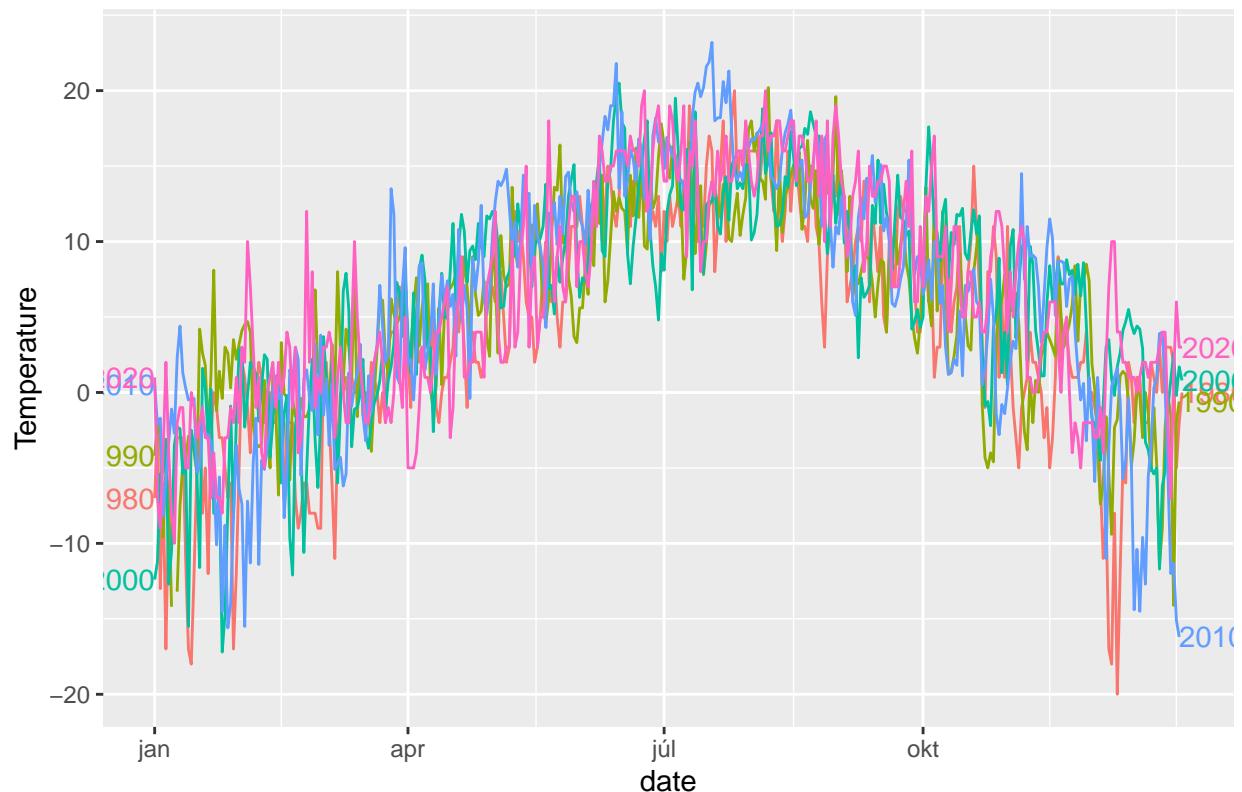
Graf sezónnosti pre teplotu - roky 1980, 1990, 2000, 2010 a 2020

Graf zobrazuje teploty pre jednotlivé sezóny počas ktorých boli merané. Vidíme, že modrá krvka pre rok 2010 v júli vystupuje nad ostatné krvky, naproti tomu červená krvka pre rok 1980 je v zimných mesiacoch nižia ako ostatné krvky.

```
all_data %>%
  dplyr::mutate(
    date = as_date(DATE)
  ) %>%
  dplyr::distinct(date, .keep_all=TRUE) %>%
  dplyr::select(date, TMP) %>%
  as_tsibble(
    index = date
  ) %>%
  tsibble::fill_gaps() %>%
  dplyr::filter(year(date) %in% c('1980', '1990', '2000', '2010', '2020')) %>%
  tsibble::fill_gaps() %>%
  gg_season(TMP, labels = "both") +
  labs(y = "Temperature",
       title = "Seasonal plot: Temperature")

## Warning: Removed 13148 row(s) containing missing values (geom_path).
## Warning: Removed 72 rows containing missing values (geom_text).
```

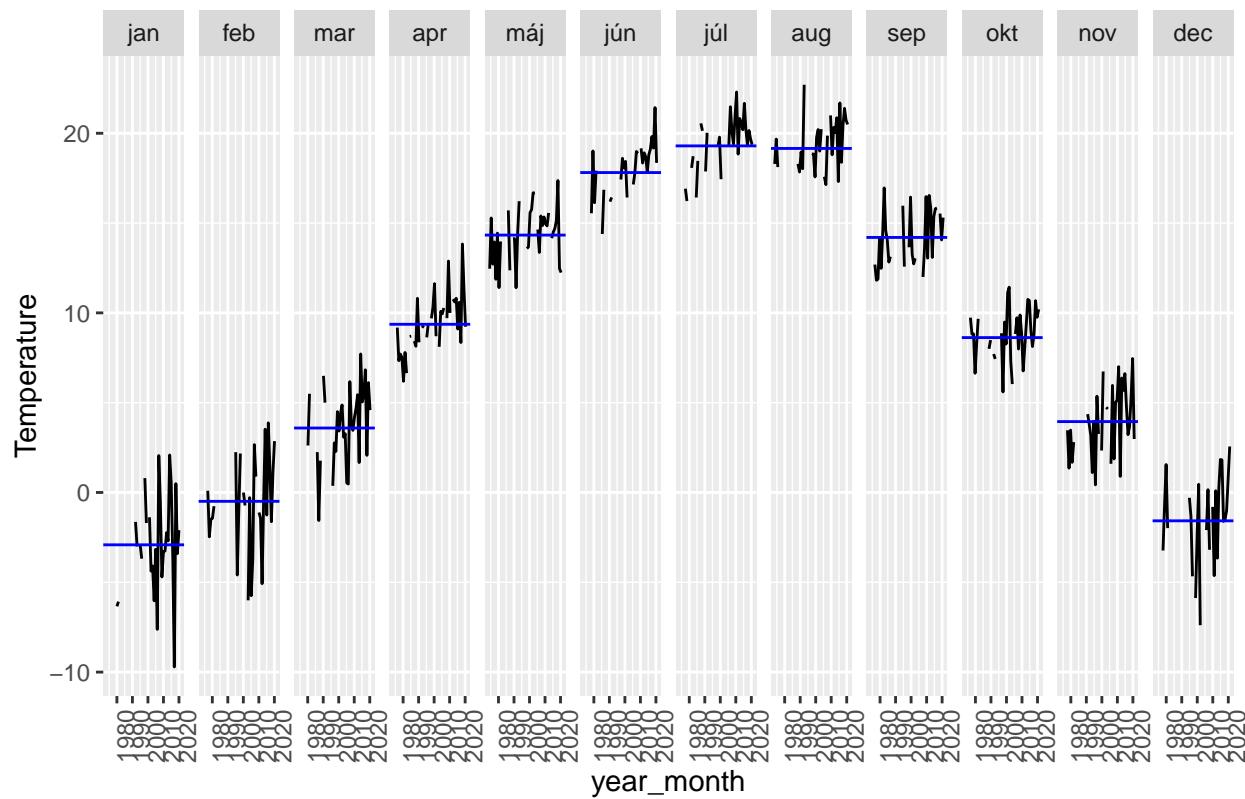
Seasonal plot: Temperature



```
all_data %>%
  dplyr::mutate(
    year_month = yearmonth(DATE)
  ) %>%
  dplyr::group_by(year_month) %>%
  dplyr::summarise(tmp = na.omit(mean(TMP))) %>%
  as_tsibble(
    index = year_month
  ) %>%
  tsibble::fill_gaps() %>%
  dplyr::filter(year(year_month)>210) %>%
  gg_subseries(tmp, period = "1 year") +
  labs(y = "Temperature",
       title = "Seasonal plot: Temperature")
```

```
## `summarise()` has grouped output by 'year_month'. You can override using the `groups` argument.
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

Seasonal plot: Temperature

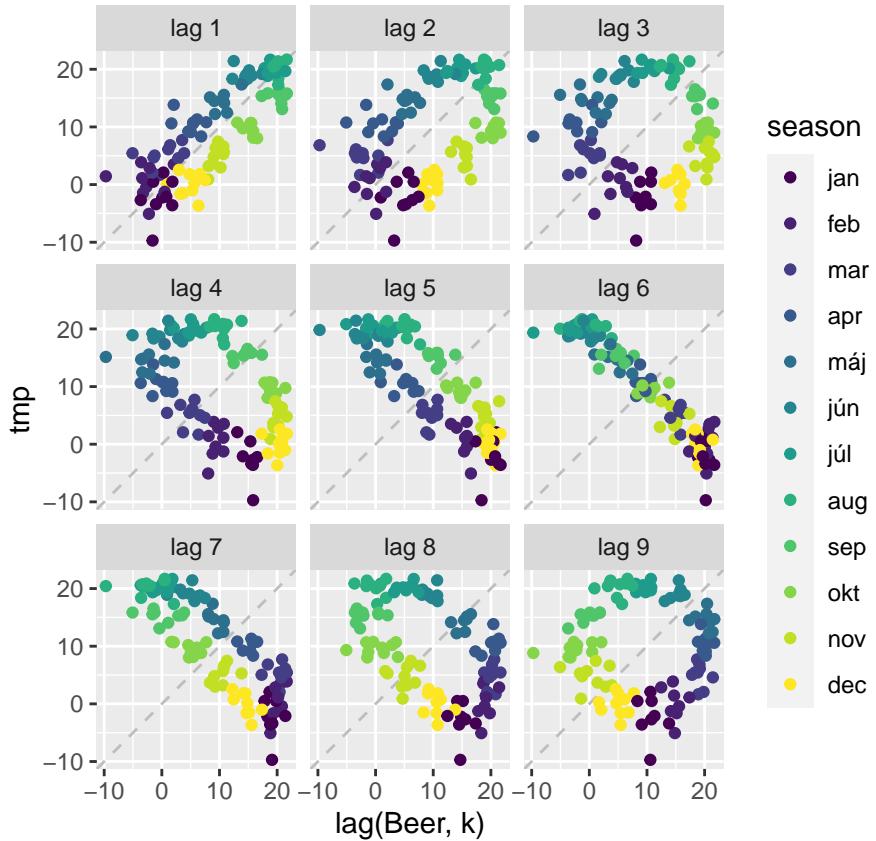


```

all_data %>%
  dplyr::mutate(
    year_month = yearmonth(DATE)
  ) %>%
  dplyr::group_by(year_month) %>%
  dplyr::summarise(tmp = na.omit(mean(TMP))) %>%
  as_tsibble(
    index = year_month
  ) %>%
  tsibble::fill_gaps() %>%
  dplyr::filter(year(year_month)>2010) %>%
  gg_lag(tmp, period = "1 year", geom = "point") +
  labs(x = "lag(Beer, k)")

## `summarise()` has grouped output by 'year_month'. You can override using the `groups` argument.
## Warning: Removed 3 rows containing missing values (gg_lag).

```



```

all_data %>%
  dplyr::mutate(
    year_month = yearmonth(DATE)
  ) %>%
  dplyr::group_by(year_month) %>%
  dplyr::summarise(tmp = na.omit(mean(TMP))) %>%
  as_tsibble(
    index = year_month
  ) %>%
  tsibble::fill_gaps() %>%
  ACF(tmp, lag_max = 9)

## `summarise()` has grouped output by 'year_month'. You can override using the `groups` argument.

## # A tsibble: 9 x 2 [1M]
##   lag      acf
##   <lag>     <dbl>
## 1 1M  0.808
## 2 2M  0.459
## 3 3M  0.0103
## 4 4M -0.434
## 5 5M -0.736
## 6 6M -0.834
## 7 7M -0.712
## 8 8M -0.388
## 9 9M -0.000154

```

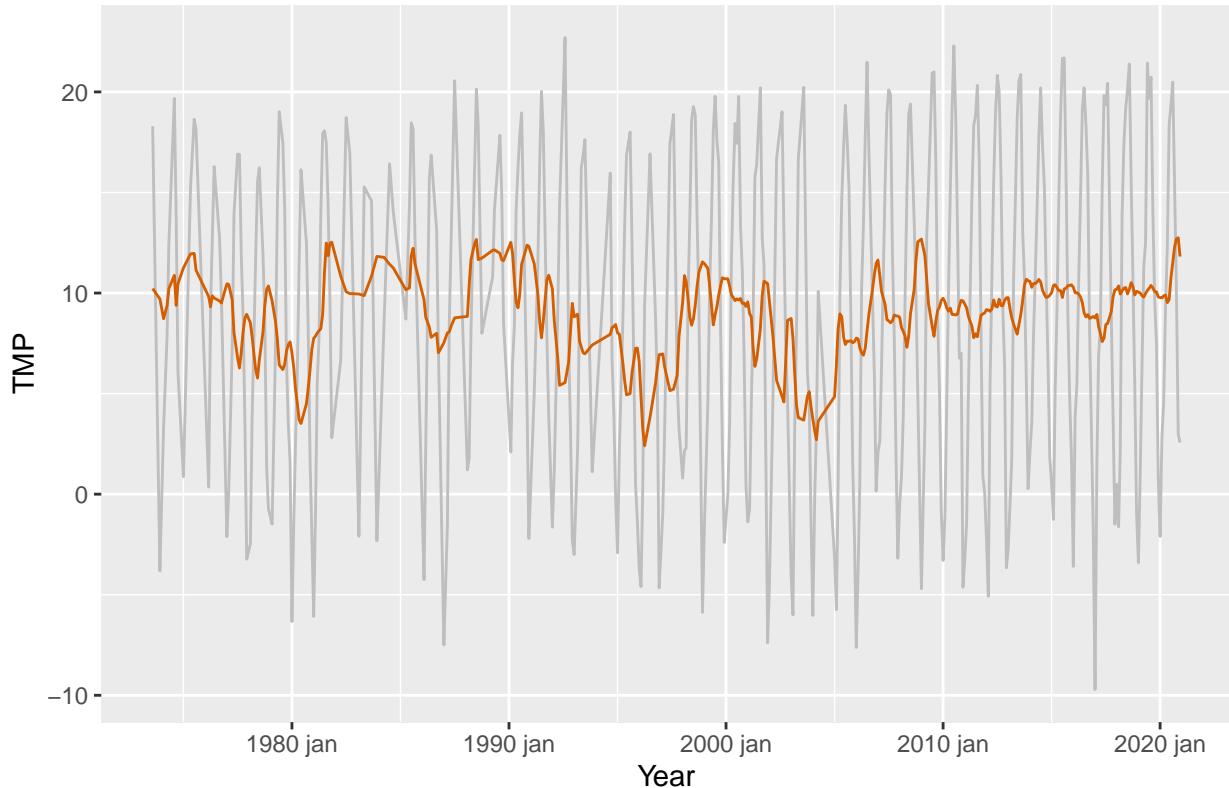
```

all_data %>%
  dplyr::mutate(
    year_month = yearmonth(DATE)
  ) %>%
  dplyr::group_by(year_month) %>%
  dplyr::summarise(tmp = na.omit(mean(TMP))) %>%
  as.data.frame %>%
  as_tsibble(
    index = year_month
  ) %>%
  dplyr::mutate(
    MA = slider::slide_dbl(tmp, mean,
                           .before = 5, .after = 6)
  ) %>%
  autoplot(tmp, colour = "gray") +
  geom_line(aes(y = MA), colour = "#D55E00") +
  labs(y = "TMP", x = "Year",
       title = "Moving average of temperature")

```

`summarise()` has grouped output by 'year_month'. You can override using the ` `.groups` argument.

Moving average of temperature



```
#components(dcmp)
```

Chýbajúce hodnoty sme nahradili pomocou funkcie **na_seadec**, ktorá z časového radu odstráni sezónnu zložku, vykoná nahradenie chýbajúcich hodnôt pomocou interpolácie a následne sezónnu zložku pridá naspäť do časového radu.

```

all_data %>%
  dplyr::mutate(
    year_month = yearmonth(DATE)
  ) %>%
  dplyr::group_by(year_month) %>%
  dplyr::summarise(tmp = na.omit(mean(TMP))) %>%
  as_tsibble(
    index = year_month
  ) %>%
  tsibble::fill_gaps() -> tsdf

## `summarise()` has grouped output by 'year_month'. You can override using the `.`groups` argument.
na_seadec(tsdf,algorithm = "interpolation", find_frequency=TRUE) %>%
  as_tsibble(
    index = year_month
  ) %>%
  model(STL(tmp )) -> m

```

Namerané hodnoty teploty sú časové rady, ktoré tvoria hodnoty zaznamenané postupne v čase (sú chronologicky usporiadane). Interval medzi jednotlivými meraniami je hodinový, teda 60 minut.

Časový rad môžeme rozdeliť na niekolko zložiek.

Trendová zložka zachytáva dlhodobé zmeny v priemernom správaní sa časového radu. Napríklad pri dlhodobom zvyšovaní hodnôt.

Sezónna zložka zachytáva periodické zmeny v časovom rade, čiže zmeny, ktoré sa opakujú pravidelne, napr. každý rok. Pri teplote je možné vidieť nárast hodnôt v letných mesiacoch a naopak pokles hodnôt v zimných mesiacoch.

Zvyšková zložka predstavuje zvyšok po odstránení sezónnej a trendovej zložky z dát.

Na nasledujúcim grafe sú zobrazené pôvodné hodnoty teploty a jednotlivé zložky časového radu teploty. Vidíme, že sezónna zložka je v priebehu rokov takmer rovnaká avšak trendová zložka ukazuje jasné stúpanie hodnôt za posledné roky.

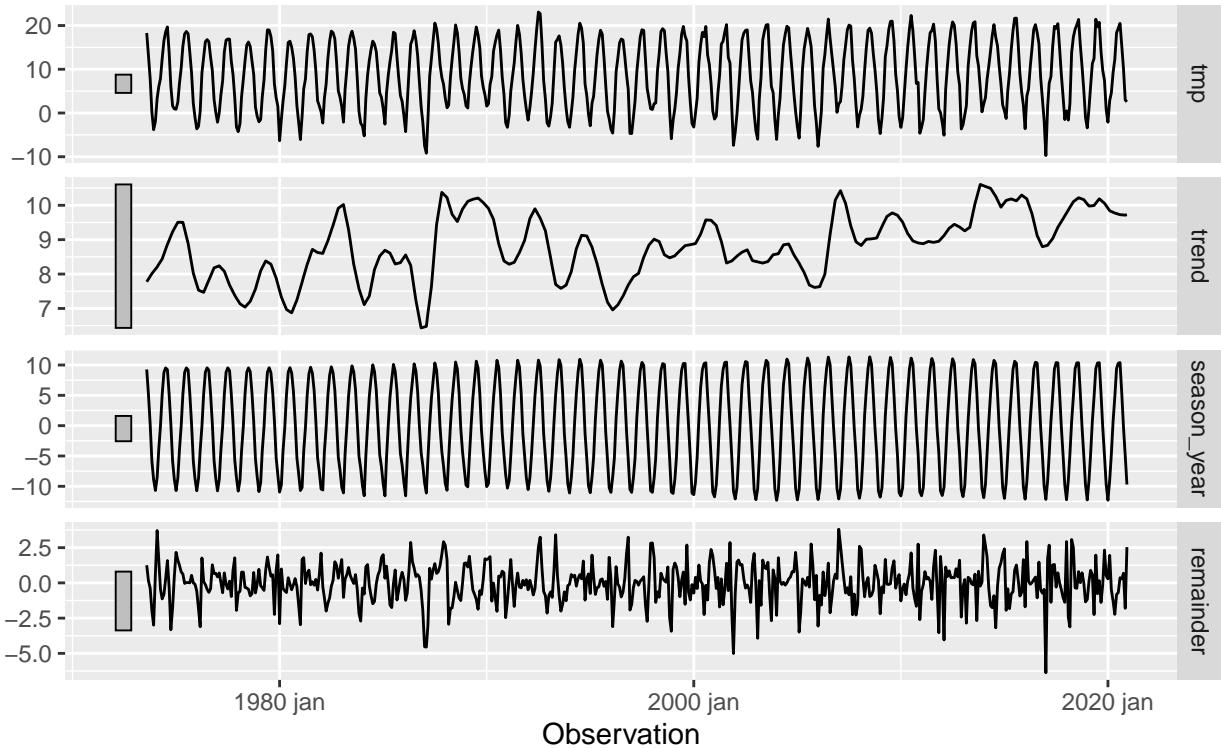
```

m %>%
  components() %>%
  autoplot() + labs(x = "Observation")

```

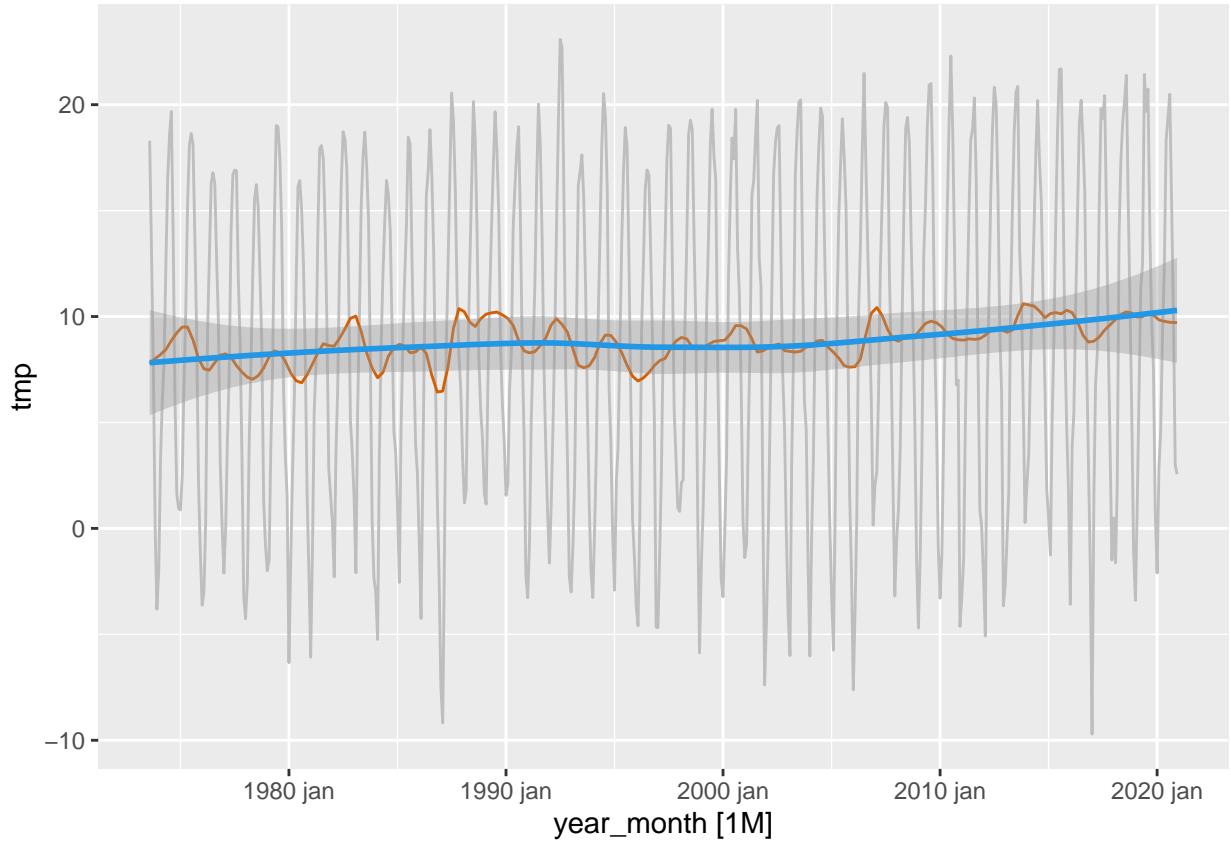
STL decomposition

tmp = trend + season_year + remainder



```
components(m) %>%
  as_tsibble() %>%
  autoplot(tmp, colour="gray") +
  geom_line(aes(y=trend), colour = "#D55E00") +
  geom_smooth(color = 4)

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



```

  labs(
    x = "Rok",
    y = "Tempeature",
    title = "Trend in avg. monthly temperature"
  ) +
  scale_y_continuous(breaks = seq(-20, 30, by = 2))

## NULL

components(m) %>%
  as_tsibble(
    index = year_month
  ) %>%
  select(year_month, trend) -> componenst_tsibble

componenst_tsibble %>%
  model(
    tsl = TSLM(trend)
  ) %>%
  forecast(h = "10 years") %>%
  autoplot(componenst_tsibble) +
  labs(title = "Predikcia - trendová zložka", x = "Rok") +
  scale_y_continuous(breaks = seq(2, 12, by = 1))

```

Predikcia – trendová zložka

