Week 07 Data visualization

Open and reproducible science: general reasons and approaches

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'readr' was built under R version 4.0.5

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Homework solutions

Task 1: data exploration

```
climatedf_comp <- read.csv(here::here("html","climate_data.csv"))
```

1.1 First look

```
head(climatedf_comp, n=3)
```

```
##   Year       Location Sunshine_duration Altitude Annual_Precipitation
## 1 1931 BaselBinningen          1594.317      316                816.0
## 2 1931 BernZollikofen          1742.500      553               1137.6
## 3 1931          Davos          1767.600     1594               1077.3
##   Annual_temperature Annual_ice_days Annual_frost_days Annual_summer_days
## 1                8.5              NA                NA                 NA
## 2                7.2              NA                NA                 NA
## 3                1.7              NA                NA                 NA
##   Annual_heat_days Annual_tropic_days Annual_precipitation_days
## 1               NA                 NA                       NA
## 2               NA                 NA                       NA
## 3               NA                 NA                       NA
```

```
summary(climatedf_comp)
```

```
##       Year        Location         Sunshine_duration    Altitude
##  Min.   :1931   Length:1170        Min.   :1046      Min.   : 273.0
##  1st Qu.:1953   Class :character   1st Qu.:1557      1st Qu.: 411.0
##  Median :1976   Mode  :character   Median :1725      Median : 485.0
##  Mean   :1976                      Mean   :1759      Mean   : 805.9
##  3rd Qu.:1998                      3rd Qu.:1937      3rd Qu.: 776.0
##  Max.   :2020                      Max.   :2741      Max.   :2501.0
##                                    NA's   :129
##  Annual_Precipitation Annual_temperature Annual_ice_days  Annual_frost_days
##  Min.   : 338.9       Min.   :-3.300     Min.   :  0.00   Min.   :  1.00
##  1st Qu.: 829.9       1st Qu.: 6.925     1st Qu.:  5.00   1st Qu.: 60.25
##  Median :1050.5       Median : 9.000     Median : 17.00   Median : 87.00
##  Mean   :1212.9       Mean   : 7.755     Mean   : 33.05   Mean   :107.99
##  3rd Qu.:1411.8       3rd Qu.:10.400     3rd Qu.: 41.00   3rd Qu.:120.00
##  Max.   :3704.2       Max.   :13.900     Max.   :218.00   Max.   :289.00
##                                          NA's   :364      NA's   :364
##  Annual_summer_days Annual_heat_days Annual_tropic_days
##  Min.   :  0.00     Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 12.25     1st Qu.: 0.000   1st Qu.: 0.000
##  Median : 38.00     Median : 2.000   Median : 0.000
##  Mean   : 36.95     Mean   : 6.257   Mean   : 1.561
##  3rd Qu.: 56.00     3rd Qu.:10.000   3rd Qu.: 1.000
##  Max.   :125.00     Max.   :56.000   Max.   :40.000
##  NA's   :364        NA's   :364      NA's   :364
##  Annual_precipitation_days
##  Min.   : 64.0
##  1st Qu.:100.0
##  Median :120.0
##  Mean   :119.8
##  3rd Qu.:136.0
##  Max.   :229.0
##  NA's   :365
```

1.2 Which has been the hottest year?

```
climatedf_comp %>%
  dplyr::group_by(Year) %>%
  dplyr::summarise(mean_temp=mean(Annual_temperature)) %>%
  dplyr::filter(mean_temp==max(mean_temp)) %>%
  dplyr::pull(Year)
```

```
## [1] 2018
```

1.3 Which has been the coldest year?

```
climatedf_comp %>%
  dplyr::group_by(Year) %>%
```

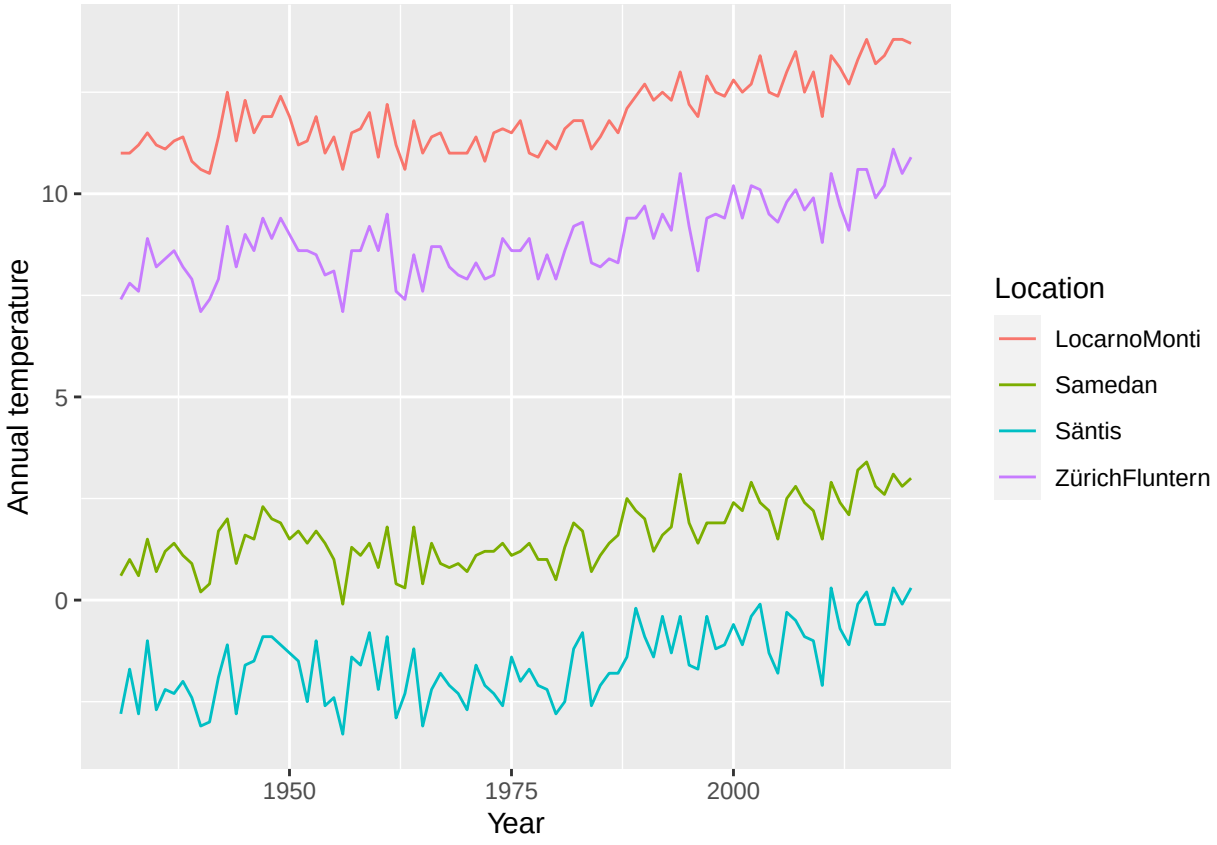
```
  dplyr::summarise(mean_temp=mean(Annual_temperature)) %>%
  dplyr::filter(mean_temp==min(mean_temp)) %>%
  dplyr::pull(Year)
```

```
## [1] 1956
```

Task 2: visualization

2.1 Association of `Annual_temperature` and `Year`

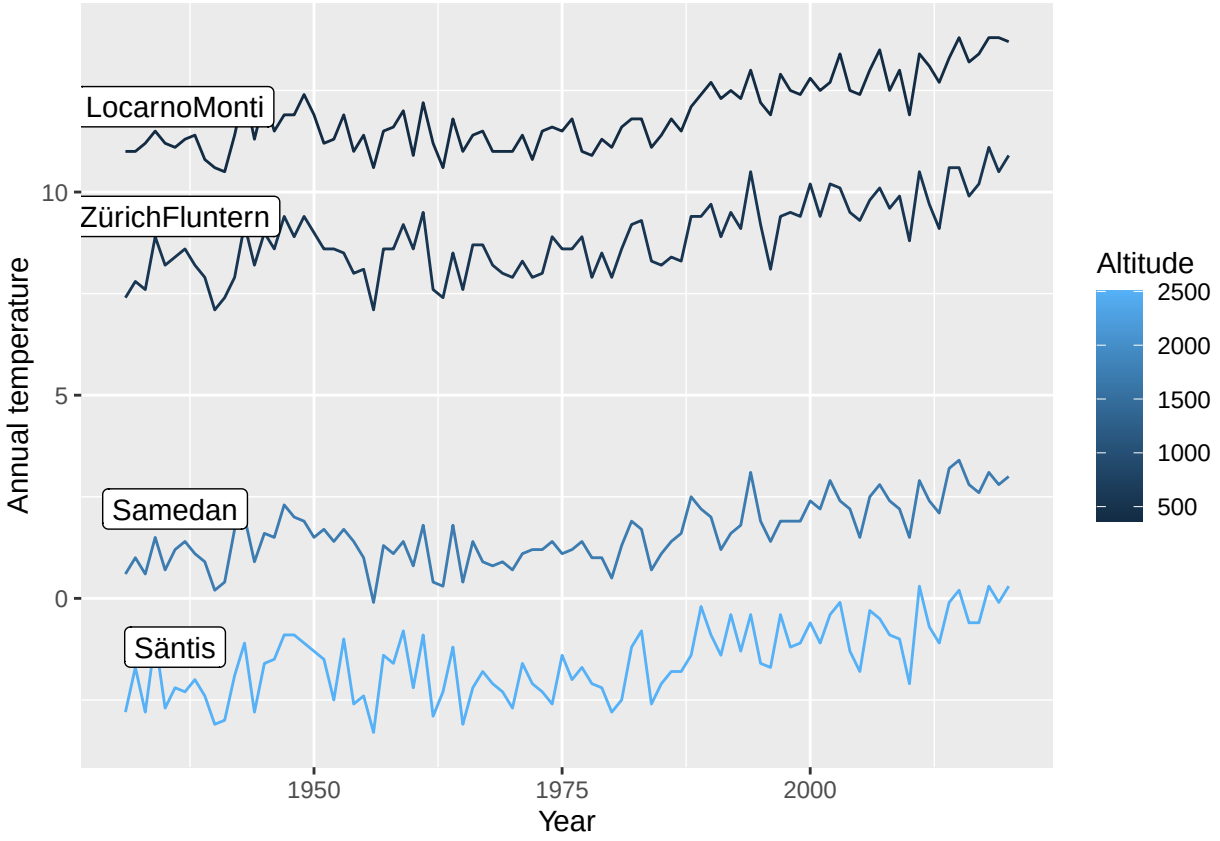
```
climatedf_comp %>%
  dplyr::filter(Location %in% c("ZürichFluntern","Säntis","Samedan","LocarnoMonti")) %>%
  ggplot() +
  geom_line(aes(Year, Annual_temperature, color=Location)) +
  labs(y="Annual temperature")
```



2.2 Add information on the altitude

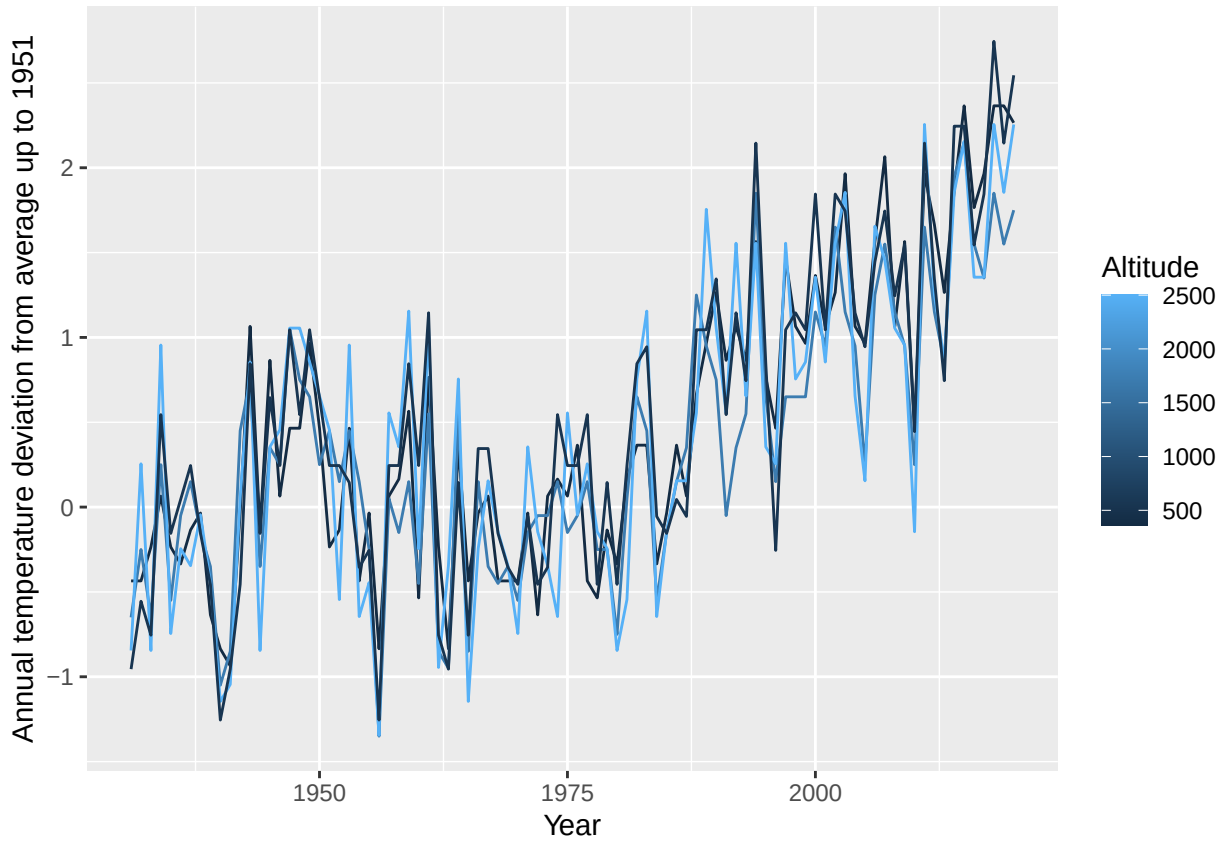
```
climatedf_comp %>%
  dplyr::filter(Location %in% c("ZürichFluntern","Säntis","Samedan","LocarnoMonti")) %>%
  ggplot() +
```

```
  geom_line(aes(Year, Annual_temperature, color=Altitude,group=Location)) +
  geom_label(aes(Year,Annual_temperature,label=Location),data = climatedf_comp %>%
  dplyr::filter(Location %in% c("ZürichFluntern","Säntis","Samedan","LocarnoMonti")) %>%
    dplyr::filter(Year==min(Year)+5),nudge_y = 1) +
  labs(y="Annual temperature")
```



2.3 Normalization

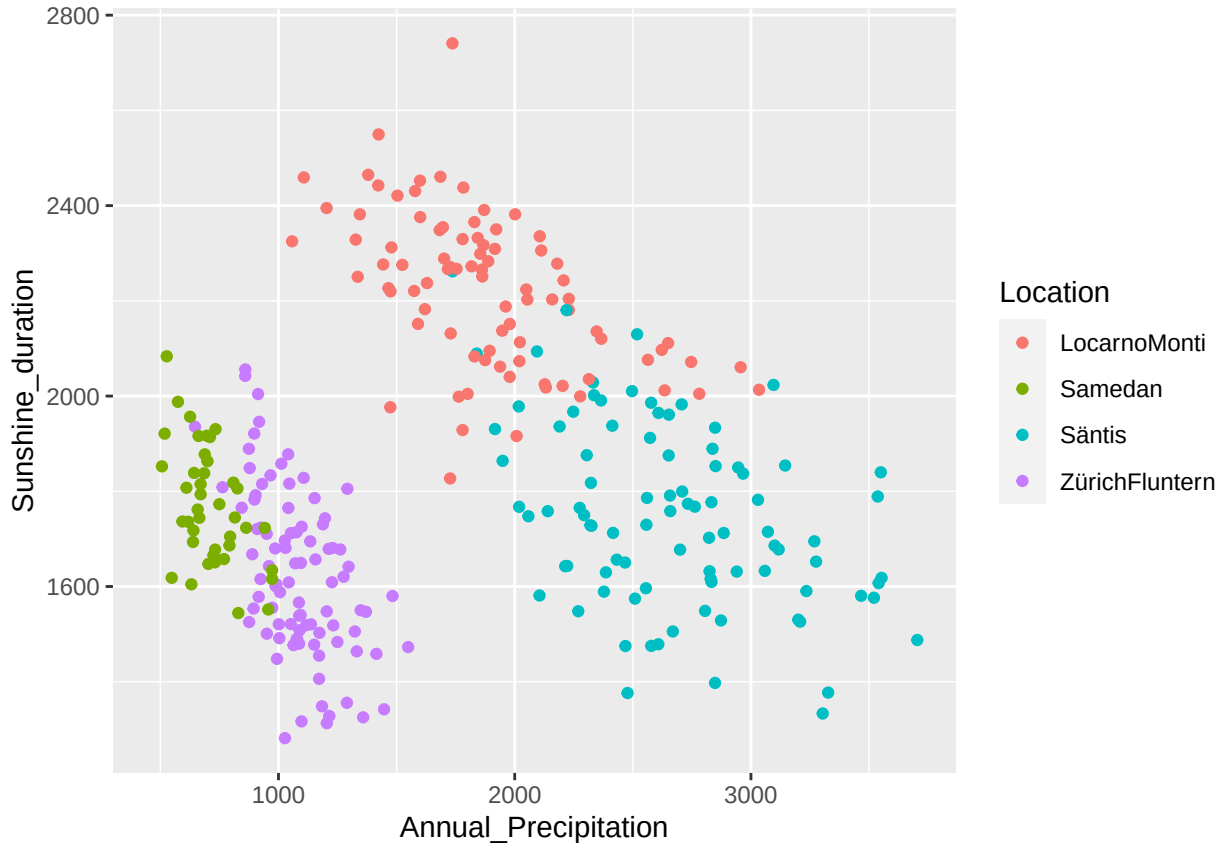
```
climatedf_comp_translated <- climatedf_comp %>%
  dplyr::filter(Location %in% c("ZürichFluntern","Säntis","Samedan","LocarnoMonti")) %>%
  dplyr::group_by(Location) %>%
  dplyr::mutate(mean_temperature = mean(Annual_temperature[Year<1951]),
                Annual_temperature = Annual_temperature-mean_temperature)

ggplot(climatedf_comp_translated) +
  geom_line(aes(Year, Annual_temperature, color=Altitude,group=Location)) +
  labs(y="Annual temperature deviation from average up to 1951")
```



2.4 Associations between `Annual_Precipitation`, and `Sunshine_duration`

```
climatedf_comp %>%
  dplyr::filter(Location %in% c("ZürichFluntern","Säntis","Samedan","LocarnoMonti")) %>%
  ggplot() +
  geom_point(aes(Annual_Precipitation, Sunshine_duration, color=Location))
```

```
## Warning: Removed 54 rows containing missing values (geom_point).
```

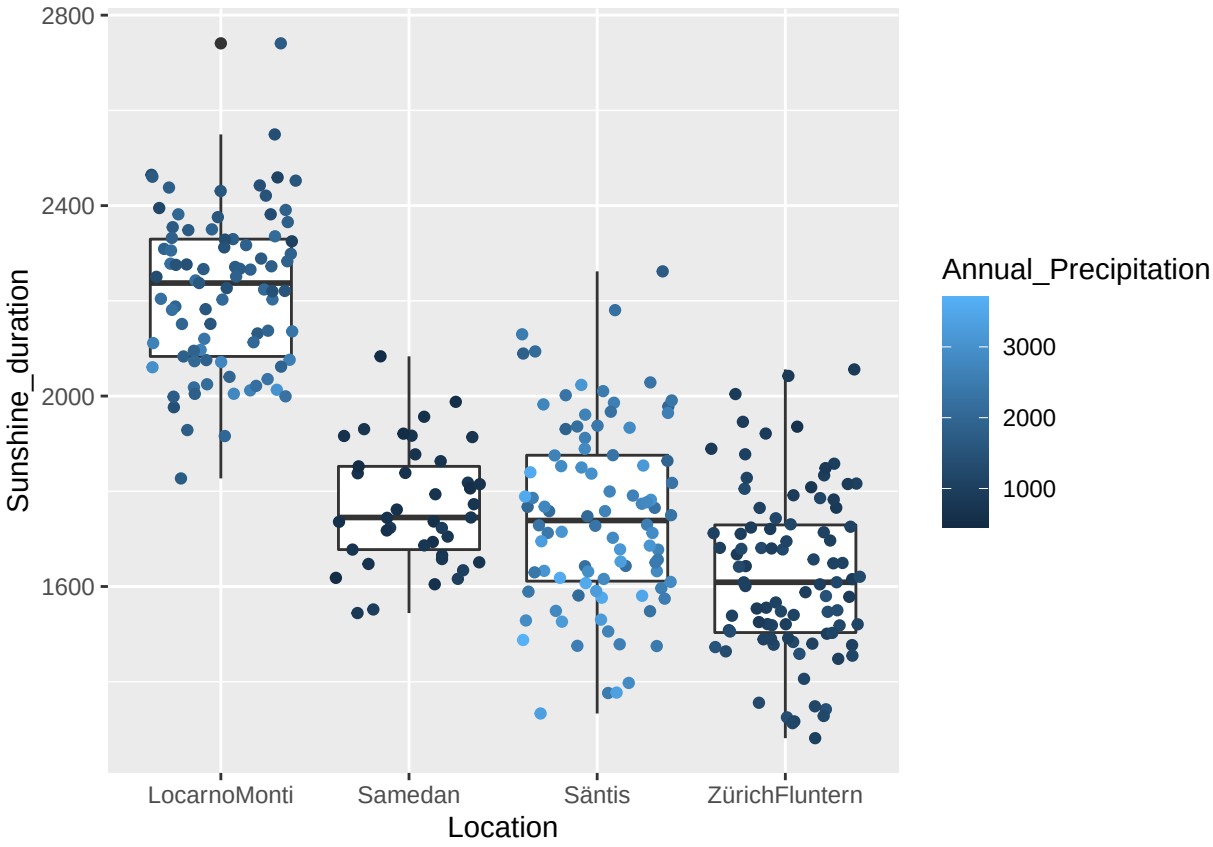



```
climatedf_comp %>%
  dplyr::filter(Location %in% c("ZürichFluntern","Säntis","Samedan","LocarnoMonti"))%>%
  ggplot() +
  geom_violin(aes(Location,Annual_Precipitation, color= Sunshine_duration)) +
  ggforce::geom_sina(aes(Location,Annual_Precipitation, color= Sunshine_duration))
```


```
climatedf_comp %>%
  dplyr::filter(Location %in% c("ZürichFluntern","Säntis","Samedan","LocarnoMonti"))%>%
  ggplot() +
  geom_boxplot(aes(Location,Sunshine_duration , color= Annual_Precipitation)) +
  geom_jitter(aes(Location,Sunshine_duration, color= Annual_Precipitation))
```

```
## Warning: Removed 54 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 54 rows containing missing values (geom_point).
```


2800
2400
2000
1600
Sunshine_duration
LocarnoMonti
Samedan
Säntis
ZürichFluntern
Location
Annual_Precipitation
3000
2000
1000