Netcdf data analysis

Guiding questions How does the mean temperature (for each year) in $\circ$C for Hamburg behave? Is there a difference to Tokio? Is there a correlation between temperature, precipitation and sun duration? How much difference is between regional values (pick any of the variables)? Why could the value be higher in some cases than in others?

Dumb Descibrtion of Data

```r
ncname <- "weather-full"
ncfname <- paste(ncname, ".nc", sep = "")

# open a netCDF file
ncin <- nc_open(ncfname)
temp <- ncvar_get(ncin, "t2m")
print(ncin)
```

```
## File weather-full.nc (NC_FORMAT_CLASSIC):
##
##      7 variables (excluding dimension variables):
##         short sf[longitude,latitude,time]
##             scale_factor: 7.3764124573405e-07
##             add_offset: 0.0241695530510217
##             _FillValue: -32767
##             missing_value: -32767
##             units: m of water equivalent
##             long_name: Snowfall
##             standard_name: lwe_thickness_of_snowfall_amount
##         short u10[longitude,latitude,time]
##             scale_factor: 0.00119632889000476
##             add_offset: -3.29942438209637
##             _FillValue: -32767
##             missing_value: -32767
##             units: m s**-1
##             long_name: 10 metre U wind component
##         short v10[longitude,latitude,time]
##             scale_factor: 0.00106014321386744
##             add_offset: 0.829670123705566
##             _FillValue: -32767
##             missing_value: -32767
##             units: m s**-1
##             long_name: 10 metre V wind component
##         short t2m[longitude,latitude,time]
##             scale_factor: 0.00203513170666401
##             add_offset: 257.975148205631
##             _FillValue: -32767
##             missing_value: -32767
##             units: K
##             long_name: 2 metre temperature
##         short e[longitude,latitude,time]
##             scale_factor: 3.68475681160163e-07
##             add_offset: -0.00823421201098244
##             _FillValue: -32767
##             missing_value: -32767
```

```
##             units: m of water equivalent
##             long_name: Evaporation
##             standard_name: lwe_thickness_of_water_evaporation_amount
##         short sund[longitude,latitude,time]
##             scale_factor: 0.659209863732776
##             add_offset: 21599.6703950681
##             _FillValue: -32767
##             missing_value: -32767
##             units: s
##             long_name: Sunshine duration
##         short tp[longitude,latitude,time]
##             scale_factor: 5.51507908704014e-06
##             add_offset: 0.180707081289331
##             _FillValue: -32767
##             missing_value: -32767
##             units: m
##             long_name: Total precipitation
##
##      3 dimensions:
##         longitude  Size:480
##             units: degrees_east
##             long_name: longitude
##         latitude  Size:241
##             units: degrees_north
##             long_name: latitude
##         time  Size:1096   *** is unlimited ***
##             units: hours since 1900-01-01 00:00:0.0
##             long_name: time
##             calendar: gregorian
##
##     2 global attributes:
##         Conventions: CF-1.0
##         history: 2015-06-03 08:02:17 GMT by grib_to_netcdf-1.13.1: grib_to_netcdf /data/data04/scrat
```

Question 1: How does the mean temperature (for each year) in $\circ$C for Hamburg (53.5511° N, 9.9937° E) behave? For this question we extracted the data for all the 3 years over Hamburg and plotted the mean tempreature of per month. The year 2016 (the years indexing is just for labeling but it doesn't reflect the actuall dates of the data measuring) seems to be the warmest year , while 2014 looks like the coldest

for easier calculations the months were all assumed to be 30 days this leads to 5 days shift for the 2nd and 3rd year, we did a test in the next figure to show the statistical difference over the mean for having those 5 days shift

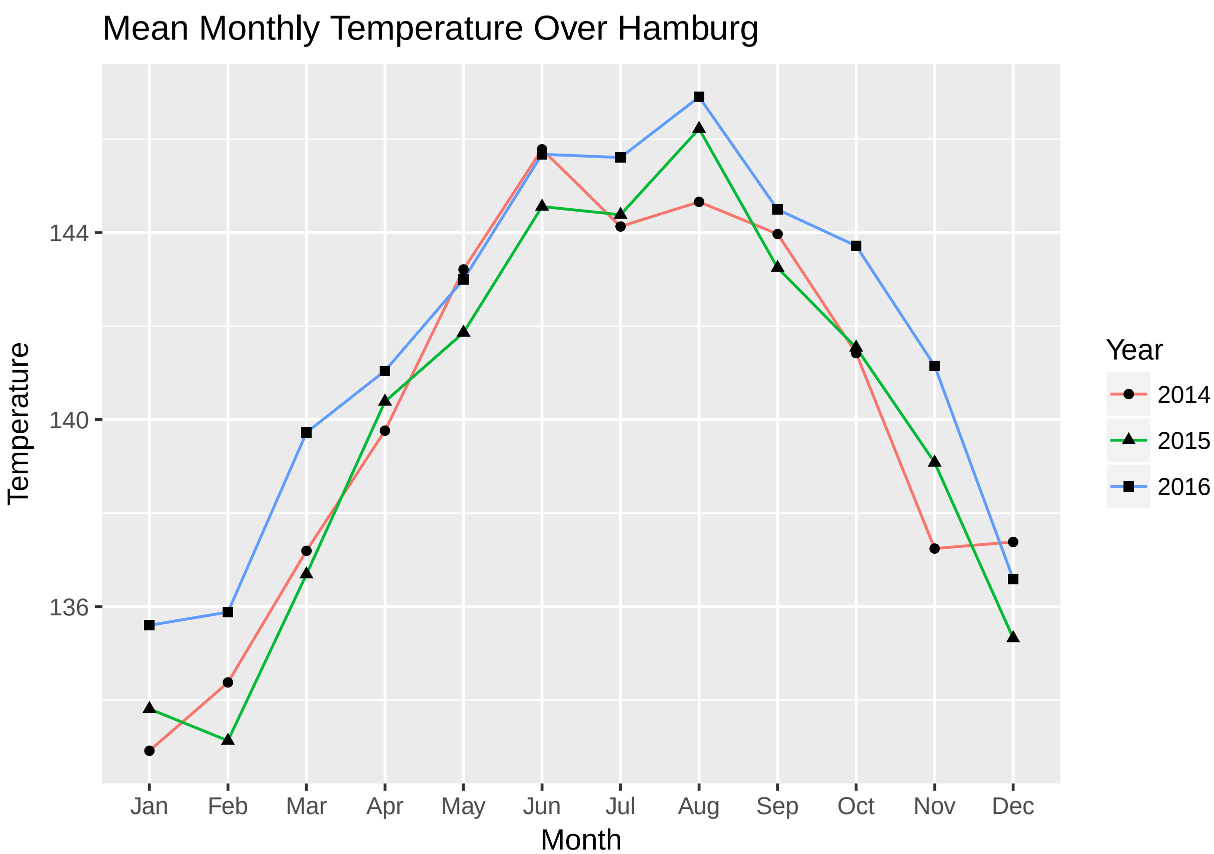
```r
library(ggplot2)
library(plyr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:plyr':
##
##     here

## The following object is masked from 'package:base':
##
##     date
```

```
# this is used as R will sort the MonthOfYear alphabitically
# if month names are used
All_Months <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul",
    "Aug", "Sep", "Oct", "Nov", "Dec")

monthly <- matrix(c(temp[14, 50, ]), nrow = 30, ncol = 36)
```

```
## Warning in matrix(c(temp[14, 50, ]), nrow = 30, ncol = 36): data length
## [1096] is not a sub-multiple or multiple of the number of rows [30]
```

```
mean_monthly <- colMeans(monthly, na.rm = TRUE)
mean_monthly <- (mean_monthly - 32) * 0.5556

dts <- seq(as.Date("2014/1/1"), by = "month", length.out = 36)
temp_hamburg <- data.frame(Temprature = c(mean_monthly), Dates = dts)
Temps_per_day <- transform(temp_hamburg, Year = format(Dates,
    "%Y"), MonthOfYear = as.numeric(format(Dates, "%m")), Temprature = Temprature)
# hist(temp_hamburg)
ggplot(Temps_per_day, aes(MonthOfYear, Temprature)) + geom_line(aes(colour = Year)) +
    geom_point(aes(shape = Year)) + scale_x_discrete(limits = All_Months) +
    labs(title = "Mean Monthly Temperature Over Hamburg") + labs(x = "Month") +
    labs(y = "Temperature")
```

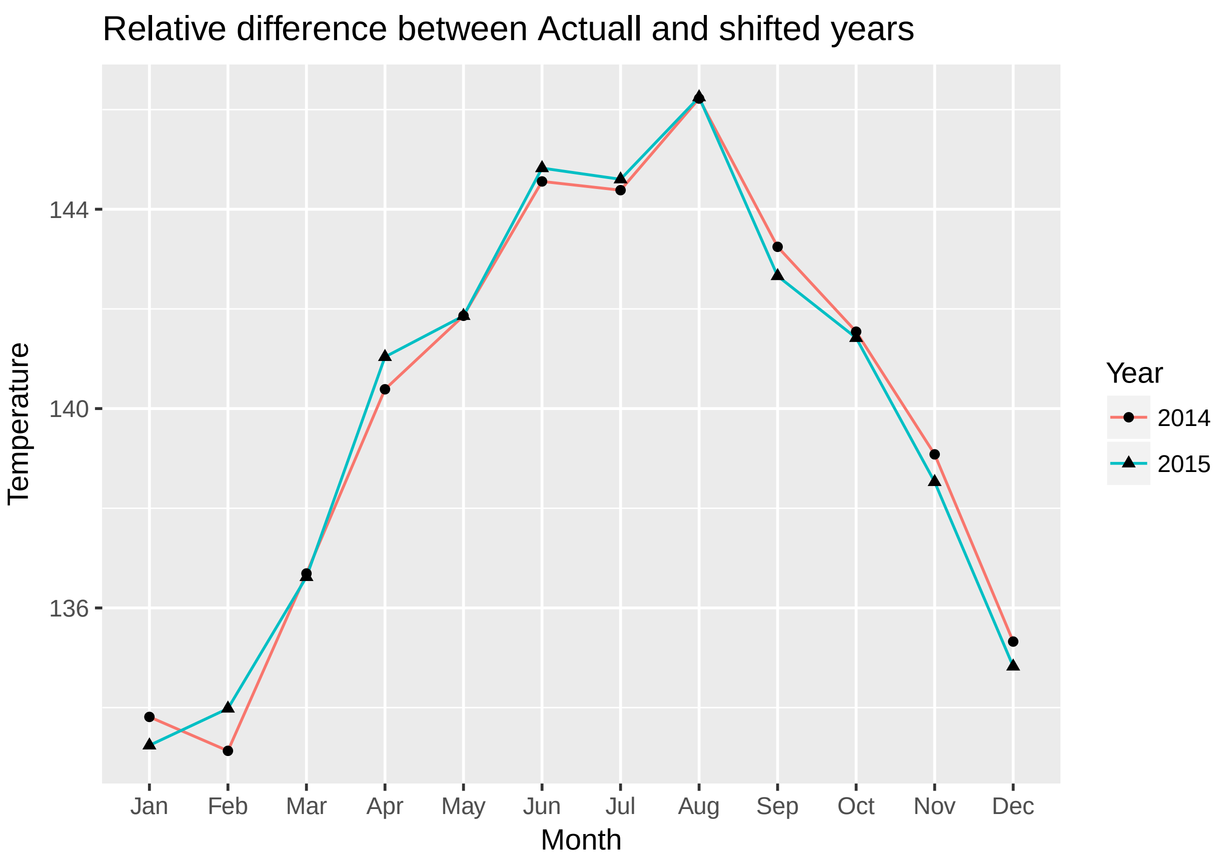


The effect of having 5 days shift in the second year. Omitting last 5 days of the actual year for simplicity. In the figures below there is actually a difference between the months specially Jan and Feb, but those differences are in the range of -0.5 to +0.5 and cannot be attributed to the differences shown in the previous figure. Thus the shifted days can be assumed to be statistically in-significant (this statement is not scientifically accurate but It's the best description available)

```r
shifte_second_year <- mean_monthly[13:24]
Actuall_second_year <- matrix(c(temp[14, 50, 365:730]), nrow = 30,
    ncol = 12)
```

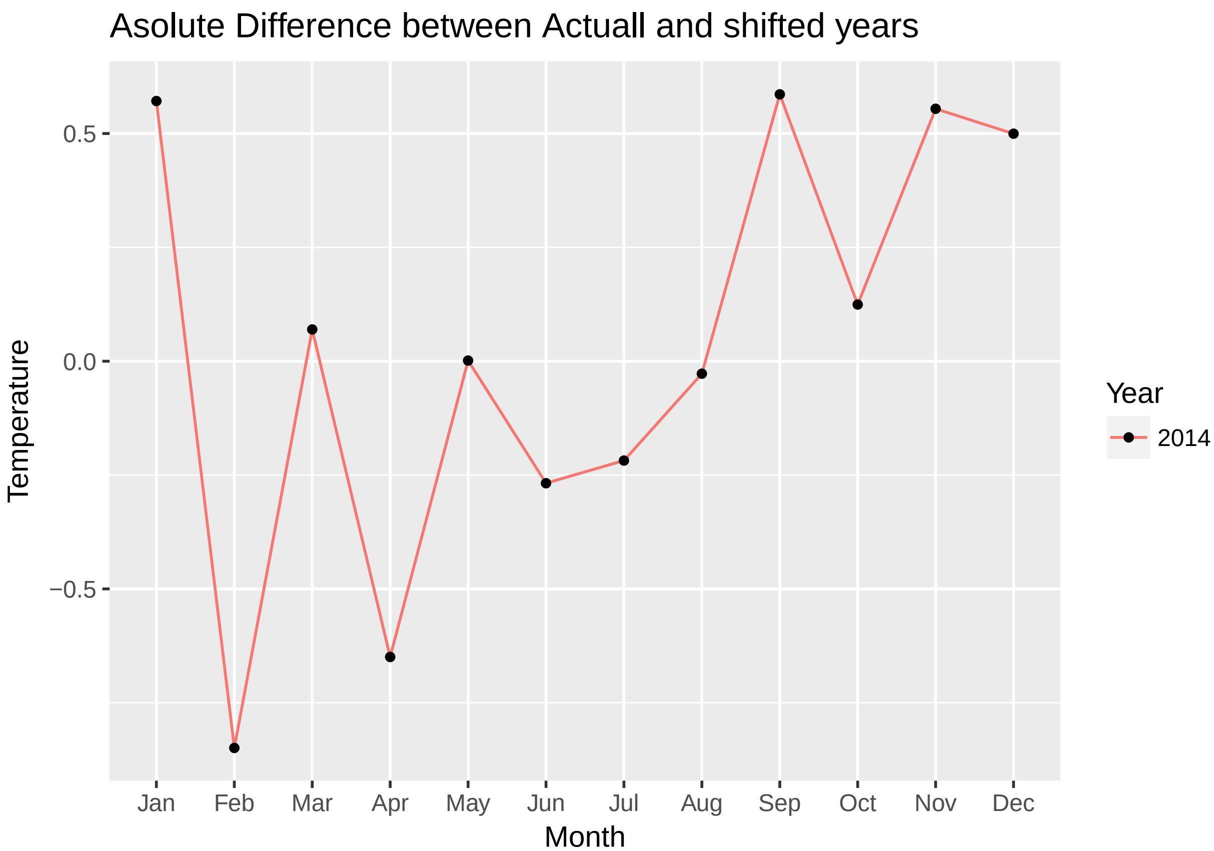
```
## Warning in matrix(c(temp[14, 50, 365:730]), nrow = 30, ncol = 12): data
## length [366] is not a sub-multiple or multiple of the number of rows [30]
```

```r
mean_monthly_second <- colMeans(Actuall_second_year, na.rm = TRUE)
mean_monthly_second <- (mean_monthly_second - 32) * 0.5556
mean_monthly_Total <- c(shifte_second_year, mean_monthly_second)

dts_shifted <- seq(as.Date("2014/1/1"), by = "month", length.out = 24)
temp_compare <- data.frame(Temprature = c(mean_monthly_Total),
    Dates = dts_shifted)
Temps_per_day <- transform(temp_compare, Year = format(Dates,
    "%Y"), MonthOfYear = as.numeric(format(Dates, "%m")), Temprature = Temprature)

ggplot(Temps_per_day, aes(MonthOfYear, Temprature)) + geom_line(aes(colour = Year)) +
    geom_point(aes(shape = Year)) + scale_x_discrete(limits = All_Months) +
    labs(title = "Relative difference between Actuall and shifted years") +
    labs(x = "Month") + labs(y = "Temperature")
```




```r
dts_shifted <- seq(as.Date("2014/1/1"), by = "month", length.out = 12)
mean_monthly_Total <- c(shifte_second_year - mean_monthly_second)
temp_compare <- data.frame(Temprature = c(mean_monthly_Total),
    Dates = dts_shifted)
Temps_per_day <- transform(temp_compare, Year = format(Dates,
    "%Y"), MonthOfYear = as.numeric(format(Dates, "%m")), Temprature = Temprature)
```

```r
ggplot(Temps_per_day, aes(MonthOfYear, Temprature)) + geom_line(aes(colour = Year)) +
    geom_point(aes(shape = Year)) + scale_x_discrete(limits = All_Months) +
    labs(title = "Asolute Difference between Actuall and shifted years") +
    labs(x = "Month") + labs(y = "Temperature")
```

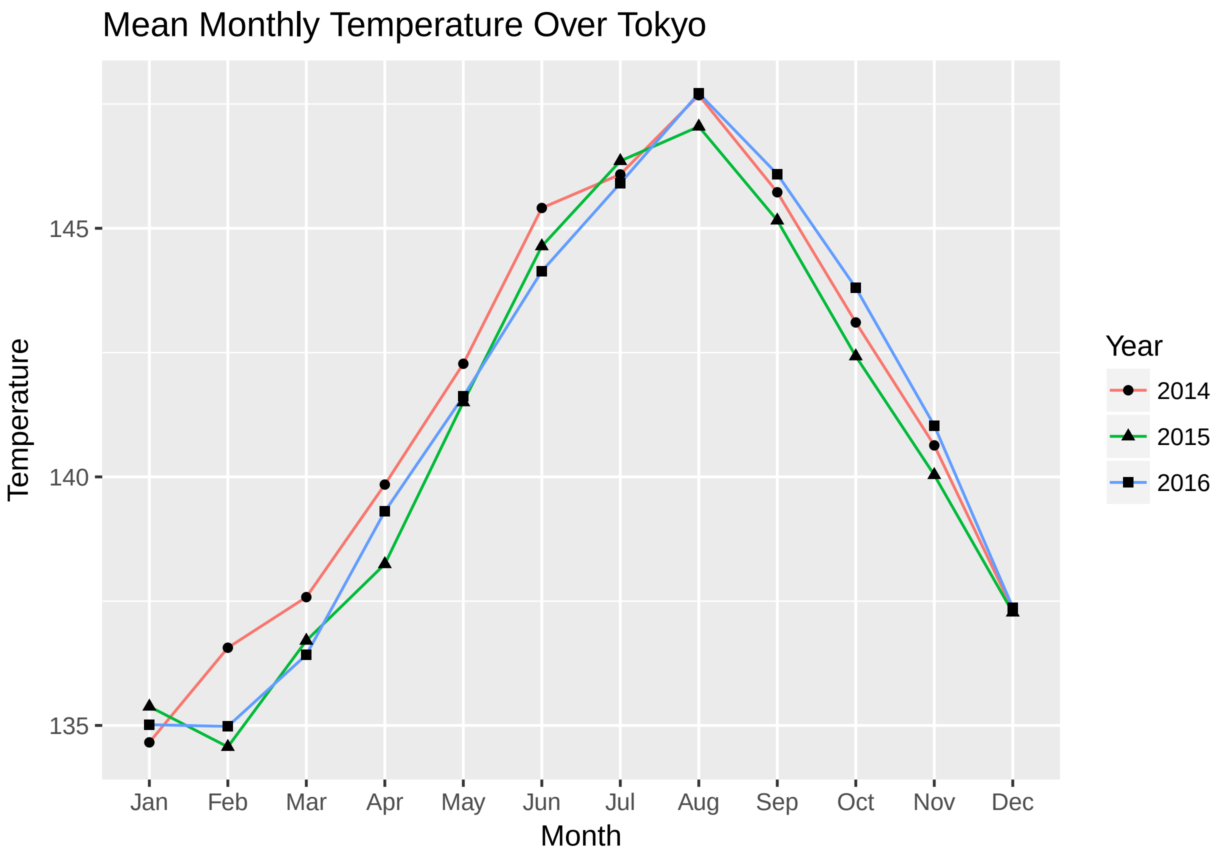


Is there a difference to Tokio (35.6895° N, 139.6917° E)? Unlike Hamburg, Tokyo seems to have a mixed weather over the years. No year is the coldest or the warmest.

```r
monthly <- matrix(c(temp[187, 74, ]), nrow = 30, ncol = 36)
```

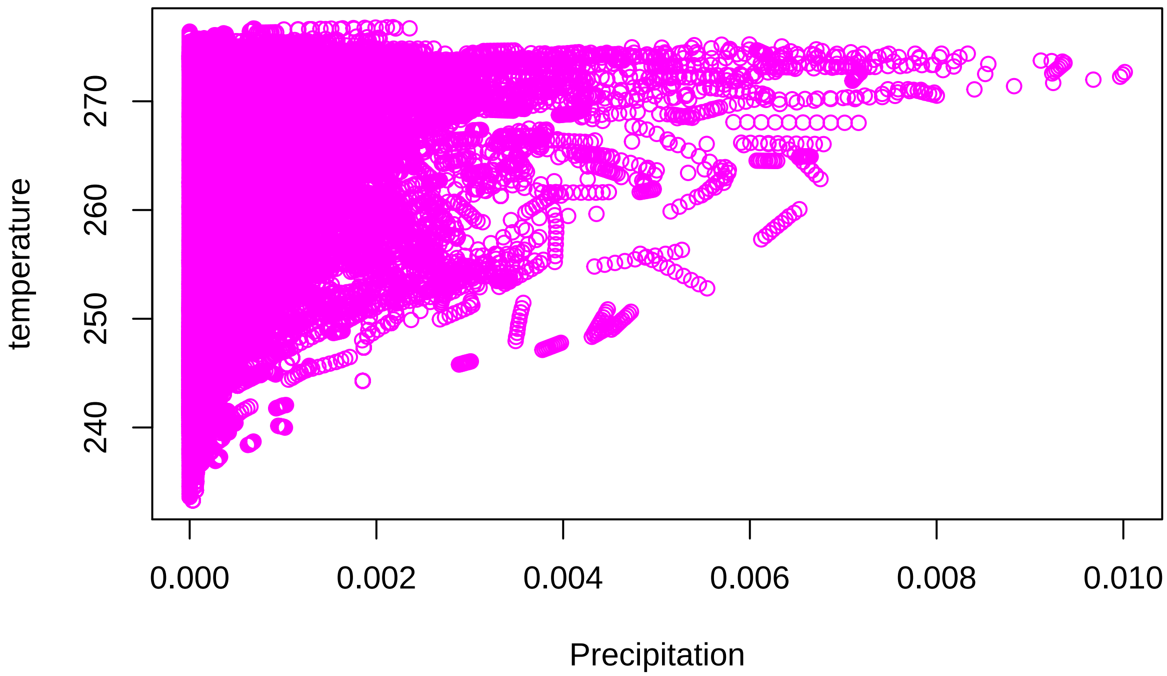
```
## Warning in matrix(c(temp[187, 74, ]), nrow = 30, ncol = 36): data length
## [1096] is not a sub-multiple or multiple of the number of rows [30]
```

```r
mean_monthly <- colMeans(monthly, na.rm = TRUE)
mean_monthly <- (mean_monthly - 32) * 0.5556
temp_hamburg <- data.frame(Temprature = c(mean_monthly), Dates = dts)
Temps_per_day <- transform(temp_hamburg, Year = format(Dates,
    "%Y"), MonthOfYear = as.numeric(format(Dates, "%m")), Temprature = Temprature)
# hist(temp_hamburg)
ggplot(Temps_per_day, aes(MonthOfYear, Temprature)) + geom_line(aes(colour = Year)) +
    geom_point(aes(shape = Year)) + scale_x_discrete(limits = All_Months) +
    labs(title = "Mean Monthly Temperature Over Tokyo") + labs(x = "Month") +
    labs(y = "Temperature")
```



Is there a correlation between temperature, precipitation and sun duration? For this question all the data for the 3 metrics have been loaded into vector, and the correlations were calculated for them. No apparent linear correlation have been detected which seems counter intuitive. For plots partial set of the actual data have been used to avoid long time in rendering.

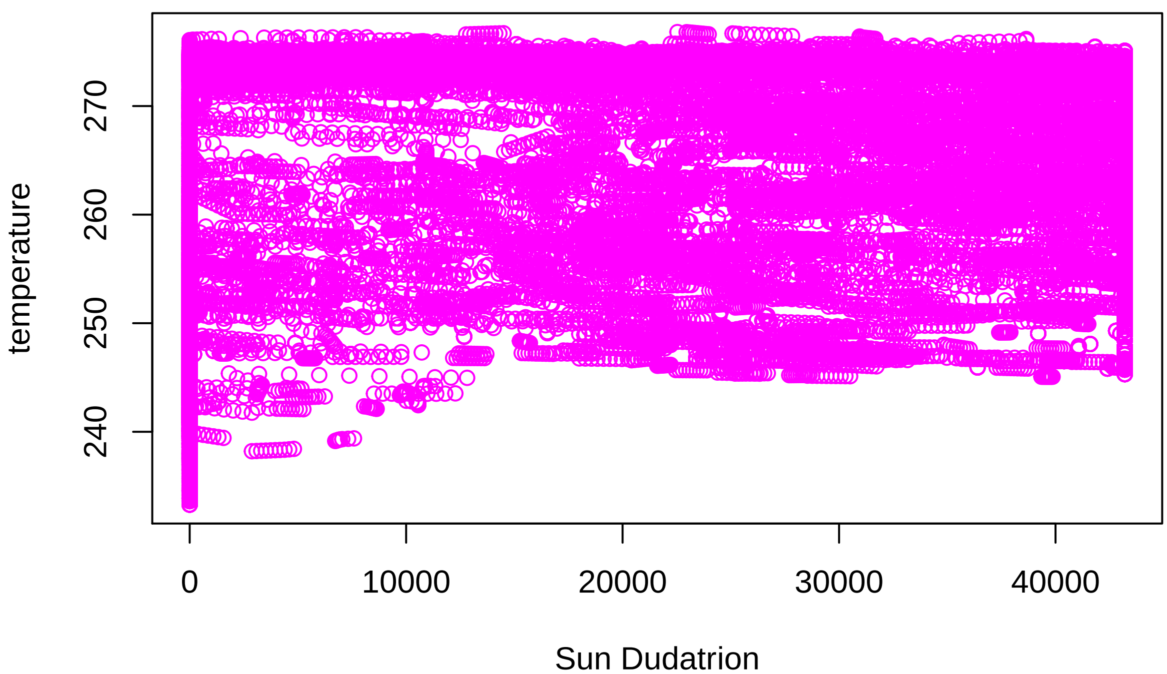
```
vector_of_temperature <- c(temp[1:10, 1:10, ])
vector_of_precipitation <- c(ncvar_get(ncin, "tp")[1:10, 1:10,
    ])
vector_of_sun_duration <- c(ncvar_get(ncin, "sund")[1:10, 1:10,
    ])


temperature <- c(temp[, , ])
precipitation <- c(ncvar_get(ncin, "tp")[, , ])
sun_duration <- c(ncvar_get(ncin, "sund")[, , ])

plot(vector_of_precipitation, vector_of_temperature, pch = 1,
    xlab = "Precipitation", ylab = "temperature", col = 14)
```




```r
cor(precipitation, temperature, method = "pearson")
```

```
## [1] 0.1892052
```

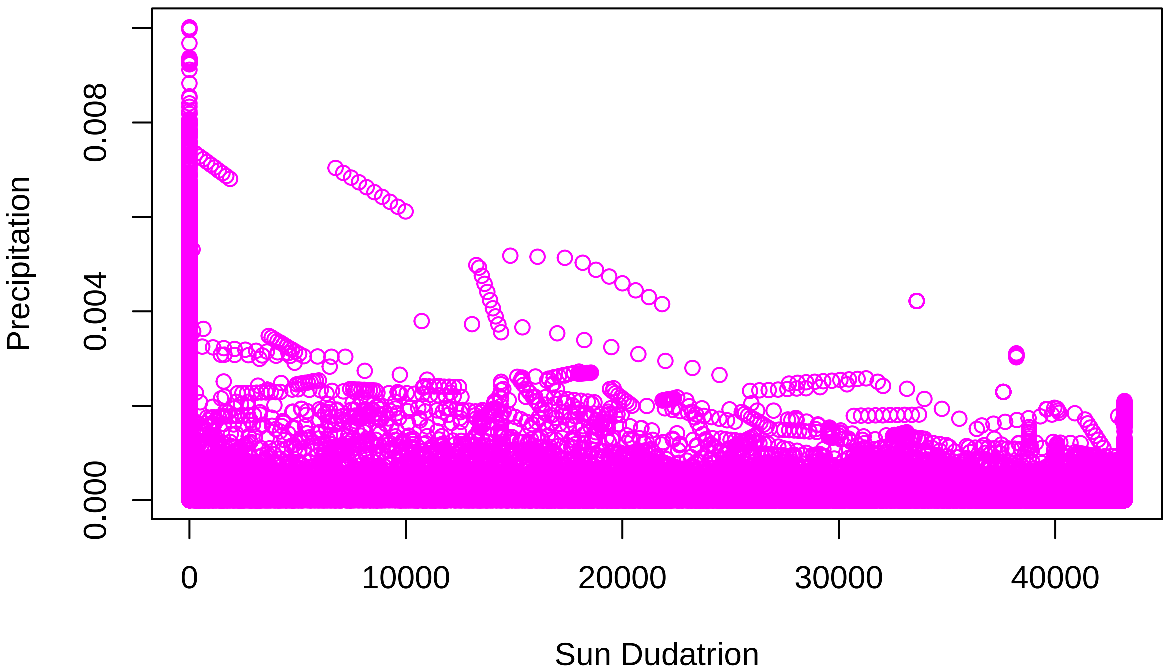
```r
plot(vector_of_sun_duration, vector_of_temperature, pch = 1,
    xlab = "Sun Dudatrion", ylab = "temperature", col = 14)
```




```r
cor(sun_duration, temperature, method = "pearson")
```

```
## [1] 0.2361691
```

```r
plot(vector_of_sun_duration, vector_of_precipitation, pch = 1,
     xlab = "Sun Dudatrion", ylab = "Precipitation", col = 14)
```

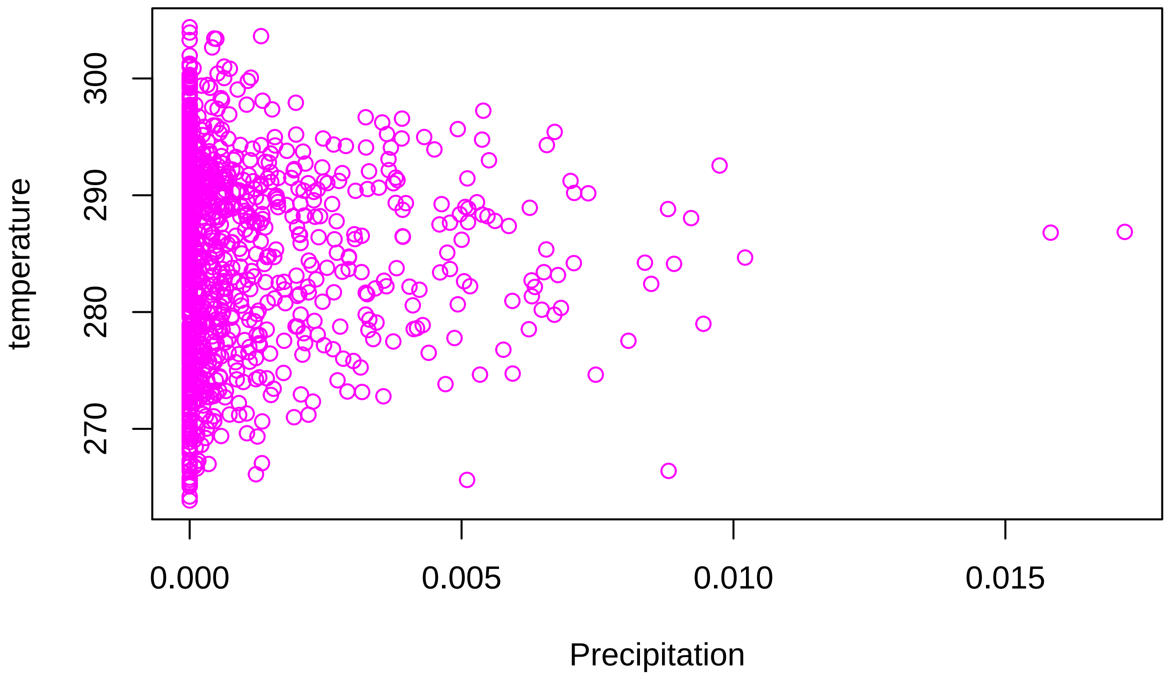



```r
cor(sun_duration, precipitation, method = "pearson")
```

```
## [1] -0.05323776
```

Is there a correlation between temperature, precipitation and sun duration? This time only testing for Hamburg and Tokyo. While unexpected but it seems that there is a positive correlation between sun duration and percipitaion in Hamburg but not in Tokyo.

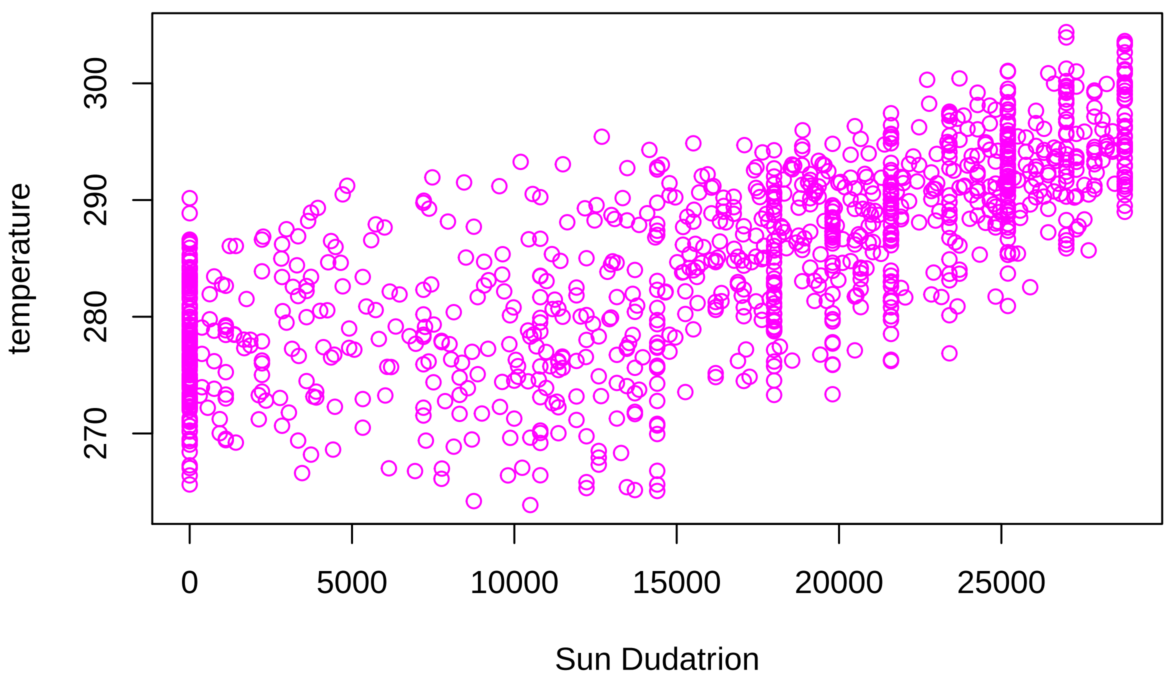
```r
vector_of_temperature <- c(temp[15, 50, ])
vector_of_precipitation <- c(ncvar_get(ncin, "tp")[15, 50, ])
vector_of_sun_duration <- c(ncvar_get(ncin, "sund")[15, 50, ])

plot(vector_of_precipitation, vector_of_temperature, pch = 1,
     xlab = "Precipitation", ylab = "temperature", col = 14)
```




```r
cor(vector_of_precipitation, vector_of_temperature, method = "pearson")
```

```
## [1] 0.003713816
```

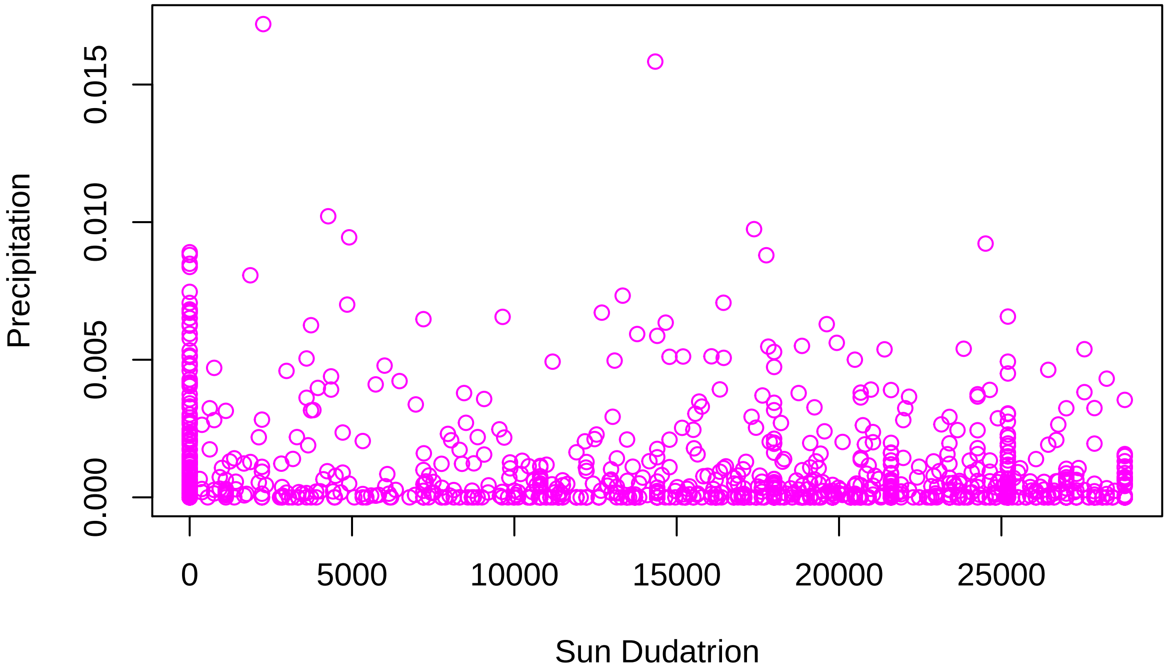
```r
plot(vector_of_sun_duration, vector_of_temperature, pch = 1,
    xlab = "Sun Dudatrion", ylab = "temperature", col = 14)
```




```r
cor(vector_of_sun_duration, vector_of_temperature, method = "pearson")
```

```
## [1] 0.7167573
```

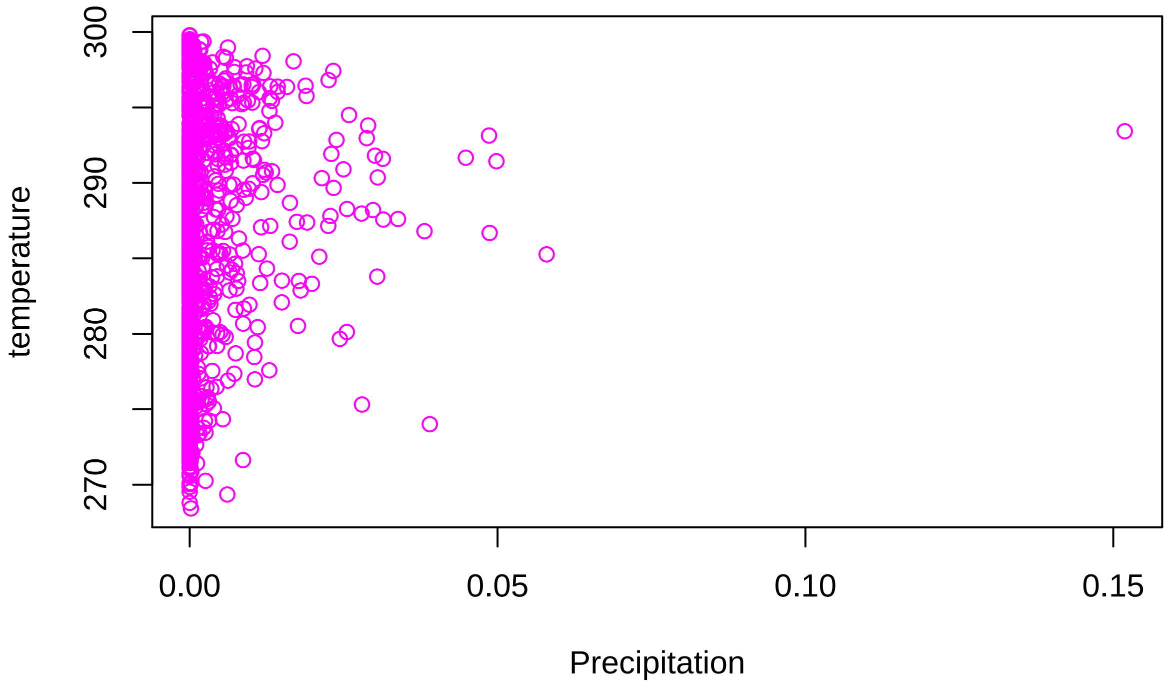
```r
plot(vector_of_sun_duration, vector_of_precipitation, pch = 1,
    xlab = "Sun Dudatrion", ylab = "Precipitation", col = 14)
```




```r
cor(vector_of_sun_duration, vector_of_precipitation, method = "pearson")
```

```
## [1] -0.2184748
```

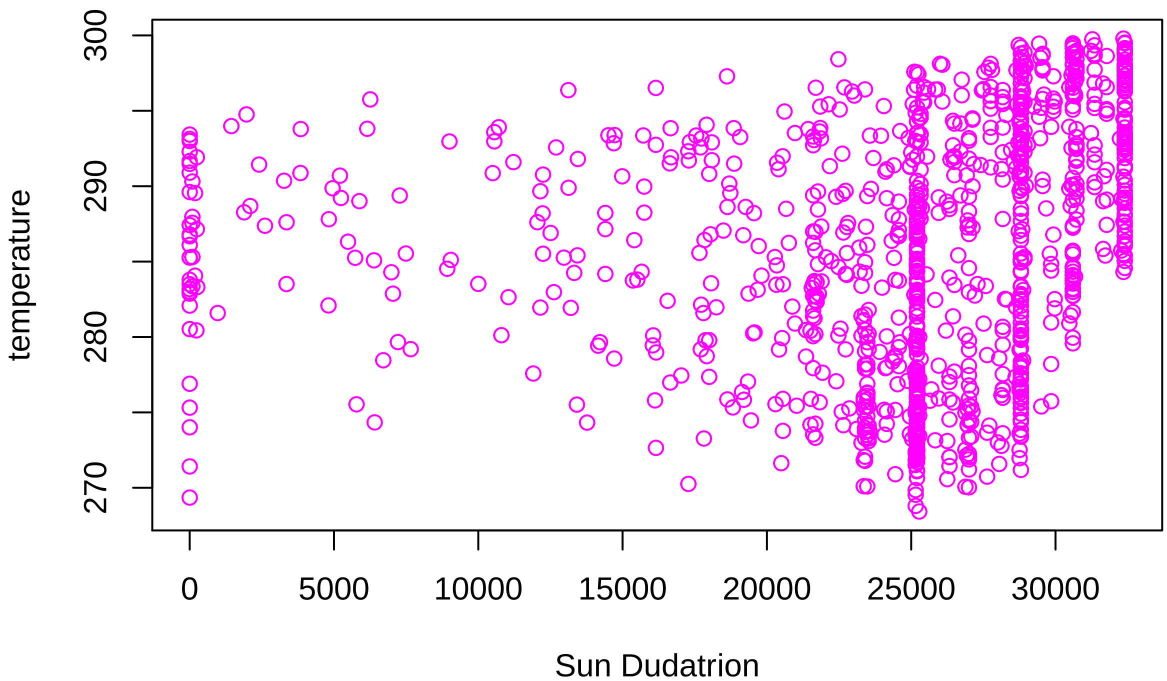
```r
vector_of_temperature <- c(temp[187, 74, ])
vector_of_precipitation <- c(ncvar_get(ncin, "tp")[187, 74, ])
vector_of_sun_duration <- c(ncvar_get(ncin, "sund")[187, 74,
    ])

plot(vector_of_precipitation, vector_of_temperature, pch = 1,
    xlab = "Precipitation", ylab = "temperature", col = 14)
```




```
cor(vector_of_precipitation, vector_of_temperature, method = "pearson")
```

```
## [1] 0.1518289
```

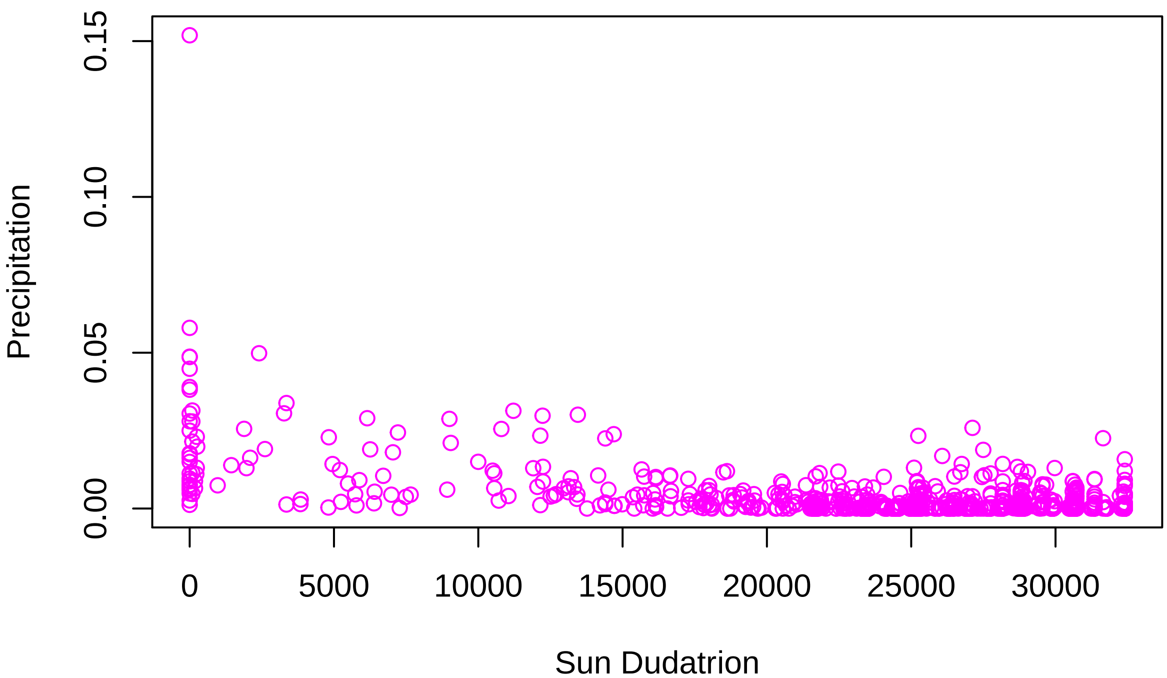
```
plot(vector_of_sun_duration, vector_of_temperature, pch = 1,
    xlab = "Sun Dudatrion", ylab = "temperature", col = 14)
```




```
cor(vector_of_sun_duration, vector_of_temperature, method = "pearson")
```

```
## [1] 0.1687673
```

```r
plot(vector_of_sun_duration, vector_of_precipitation, pch = 1,
    xlab = "Sun Dudatrion", ylab = "Precipitation", col = 14)
```




```r
cor(vector_of_sun_duration, vector_of_precipitation, method = "pearson")
```

```
## [1] -0.5314731
```