Day12 exercise solutions

Ali Movasati, Isabelle Caroline Rose Cretton, Tristan Koning

Nov. 15th, 2024

```
# Set global code chunk options
knitr::opts_chunk$set(
  echo = TRUE,
  warning = FALSE,
  message = FALSE,
  fig.width = 10,
  fig.height = 6
)
```

```
# load required libraries
library("skimr")
library("extremefit")
library("extRemes")

`%notin%` <- Negate(`%in%`)
```

Problem 1

(a)

```
data("dataWind")
str(dataWind)
```

```
## 'data.frame':    10903 obs. of  4 variables:
##  $ Year : num  1976 1976 1976 1976 1976 ...
##  $ Month: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Day  : num  2 3 4 5 6 7 8 9 10 11 ...
##  $ Speed: num  18 8.1 5.9 5.3 5.5 9.8 8.2 6.1 11.9 10.1 ...
```

```
head(dataWind)
```

```
##   Year Month Day Speed
## 1 1976     1   2  18.0
## 2 1976     1   3   8.1
## 3 1976     1   4   5.9
## 4 1976     1   5   5.3
## 5 1976     1   6   5.5
## 6 1976     1   7   9.8
```

```
skimr::skim(dataWind)
```

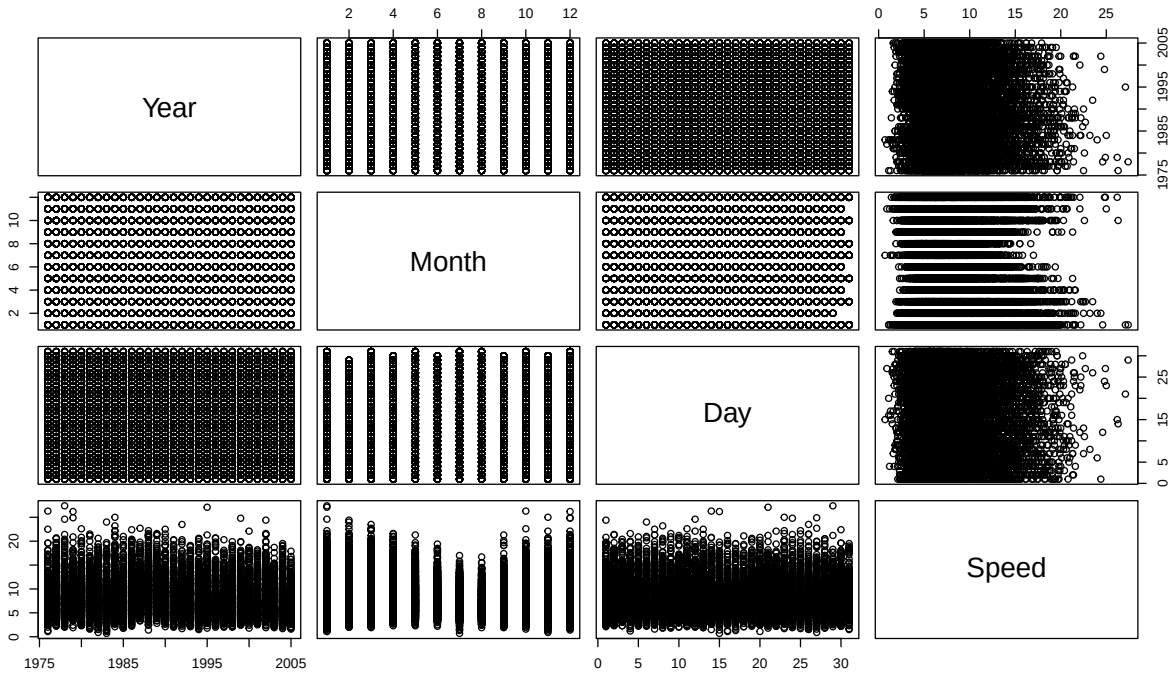
Table 1: Data summary

Name	dataWind
Number of rows	10903
Number of columns	4
Column type frequency:	
numeric	4
Group variables	None

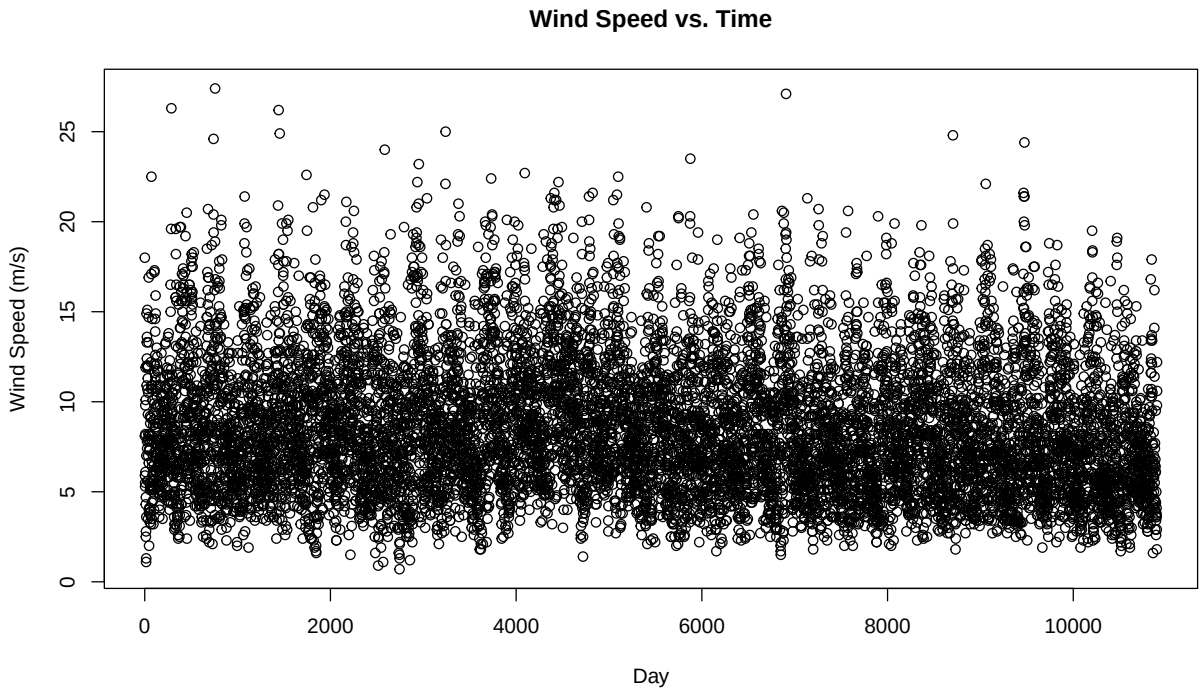
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Year	0	1	1990.51	8.67	1976.0	1983.0	1991.0	1998.0	2005.0	
Month	0	1	6.53	3.45	1.0	4.0	7.0	10.0	12.0	
Day	0	1	15.73	8.80	1.0	8.0	16.0	23.0	31.0	
Speed	6	1	8.55	3.75	0.7	5.7	7.9	10.8	27.4	

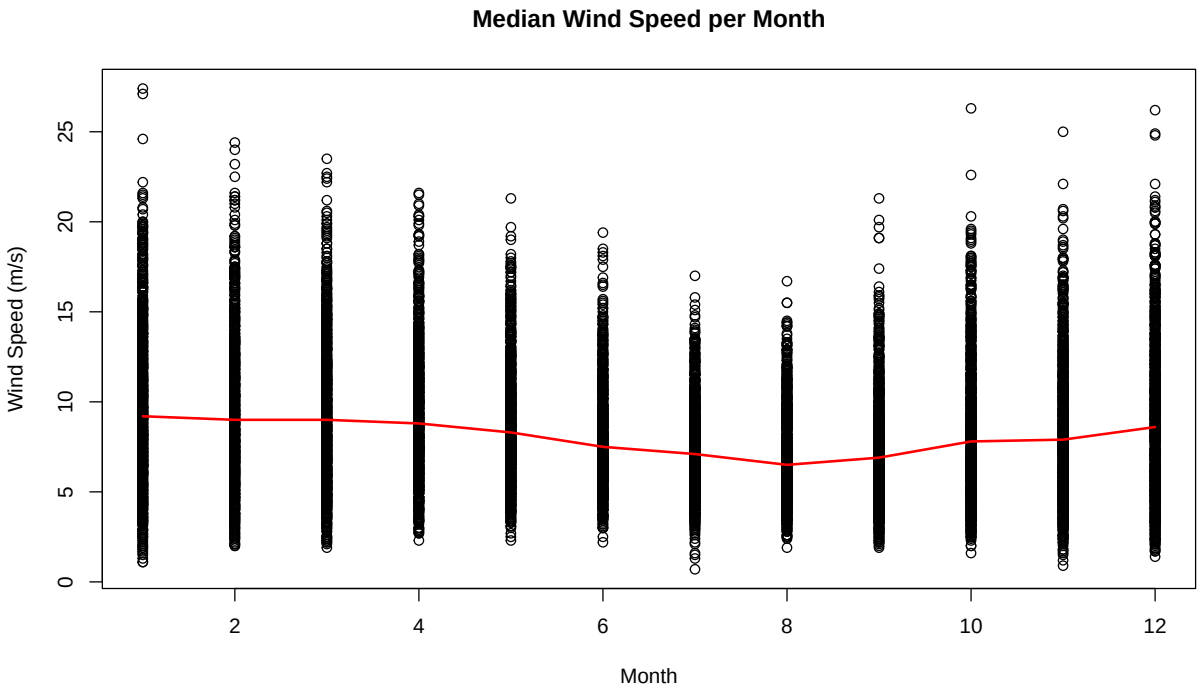
```
# Plot wind speed vs day
pairs(dataWind)
```



```r
plot(dataWind$Speed, xlab="Day", ylab="Wind Speed (m/s)", main="Wind Speed vs. Time")
```

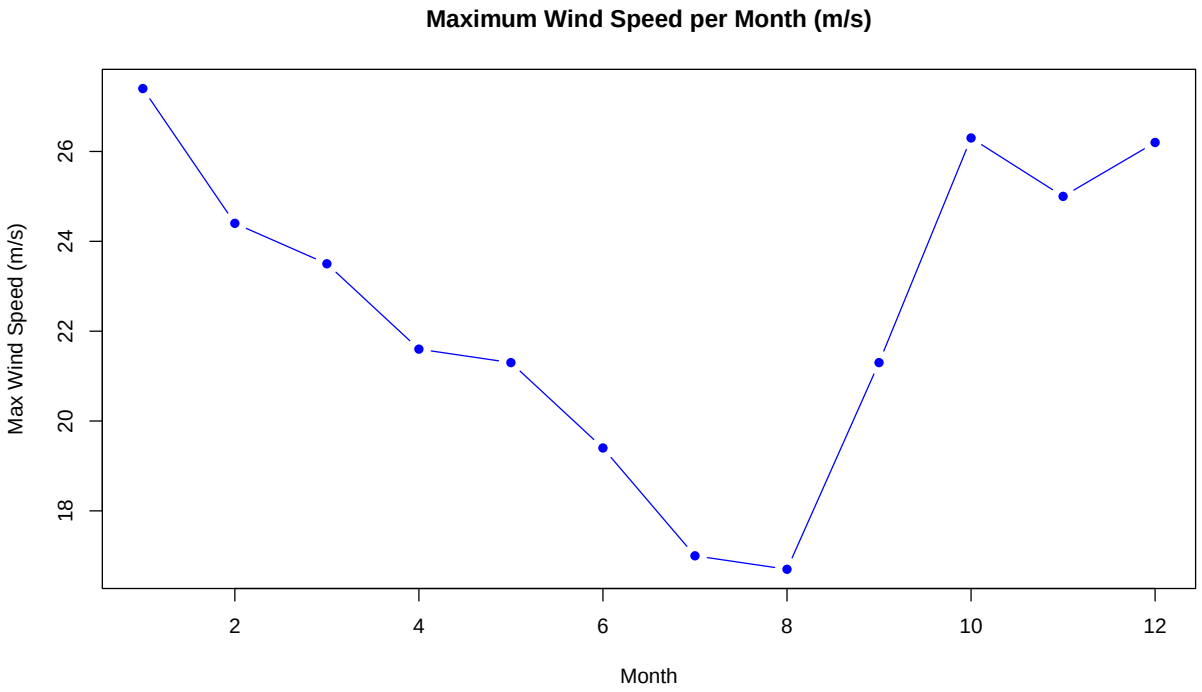



```r
monthly_med <- aggregate(Speed~Month, data=dataWind, median)
plot(Speed~Month, xlab="Month", ylab="Wind Speed (m/s)", main="Median Wind Speed per Month", data=dataW
lines(monthly_med$Month, monthly_med$Speed, col="red", lwd=2)
```




```r
max_monthly <- aggregate(Speed~Month, data=dataWind, max)

plot(x=max_monthly$Month, y=max_monthly$Speed,
     xlab="Month",
     ylab="Max Wind Speed (m/s)",
     main="Maximum Wind Speed per Month (m/s)",
     pch=16,
     type="b",
     col="blue")
```



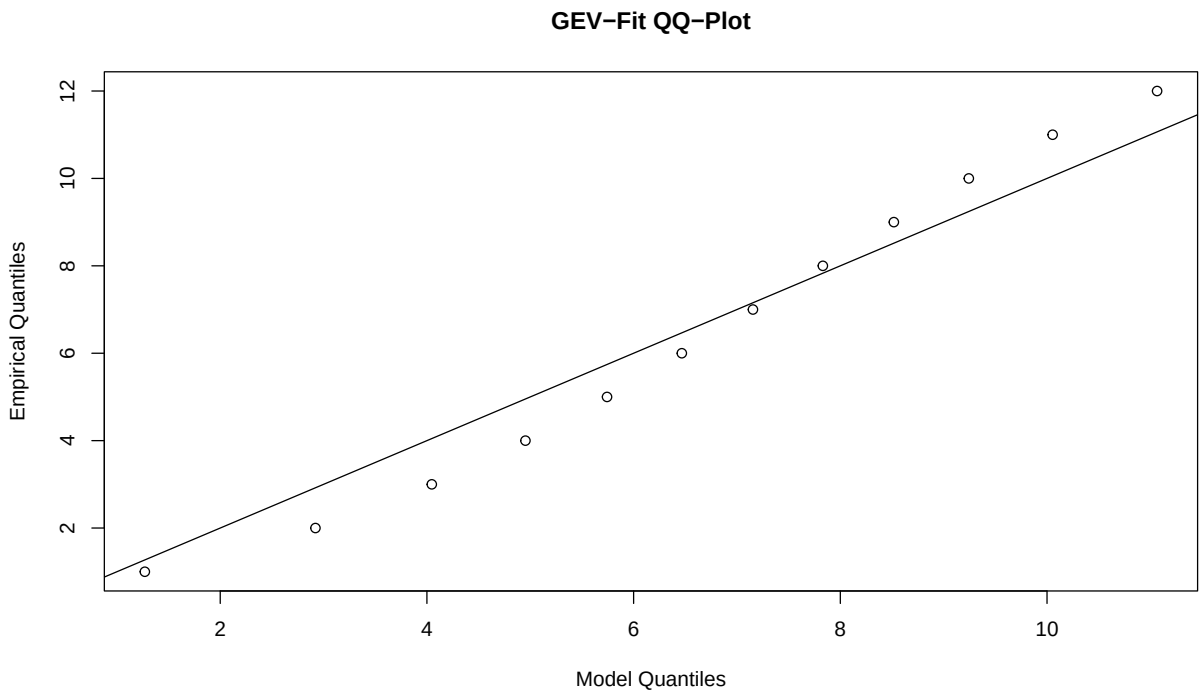

```r
gev <- fevd(max_monthly$Month)
summary(gev)
```

```
##
## fevd(x = max_monthly$Month)
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  31.5707
##
##
##  Estimated parameters:
##   location      scale      shape
##  5.5779814  3.6603616 -0.4563442
##
##  Standard Error Estimates:
##  location     scale     shape
## 1.2597746 1.0640837 0.3461606
##
##  Estimated parameter covariance matrix.
##            location      scale      shape
## location  1.5870321  0.1025138 -0.2329187
## scale     0.1025138  1.1322741 -0.2813401
## shape    -0.2329187 -0.2813401  0.1198271
##
##  AIC = 69.1414
##
##  BIC = 70.59612
```

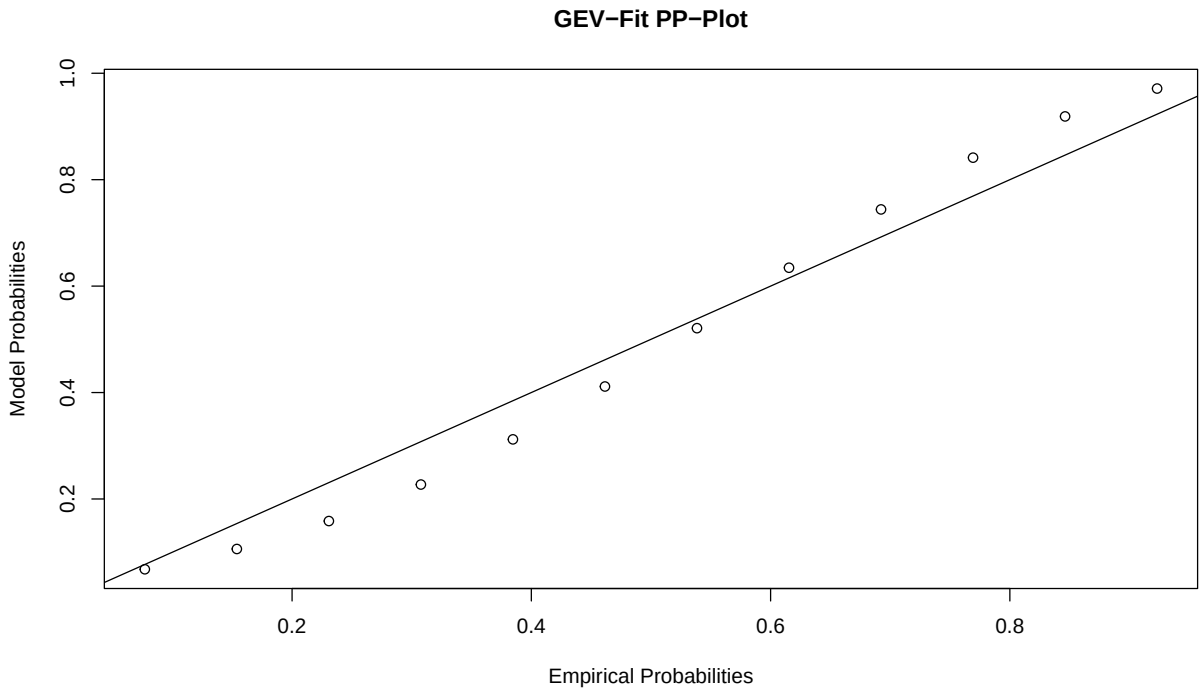
```r
ci(gev, type="parameter")
```

```
## fevd(x = max_monthly$Month)
##
## [1] "Normal Approx."
##
##          95% lower CI   Estimate 95% upper CI
## location     3.108868  5.5779814    8.0470943
## scale        1.574796  3.6603616    5.7459273
## shape       -1.134806 -0.4563442    0.2221181
```

```r
plot(gev, type="qq", main="GEV-Fit QQ-Plot")
```




```r
plot(gev, type="prob", main="GEV-Fit PP-Plot")
```



Looking at both the QQ- and PP-Plot, it is safe to assume that the model we have created is not suited for the real data, as in both plots we can observe quite large deviations.

```
gpd <- fevd(na.omit(dataWind$Speed), threshold=20, type="GP", time.units="months")
summary(gpd)
```

```
##
## fevd(x = na.omit(dataWind$Speed), threshold = 20, type = "GP",
##     time.units = "months")
##
## [1] "Estimation Method used: MLE"
##
##
##  Negative Log-Likelihood Value:  105.6396
##
##
##  Estimated parameters:
##      scale      shape
##  1.8965652 -0.0394441
##
##  Standard Error Estimates:
##     scale     shape
## 0.3551409 0.1411950
##
##  Estimated parameter covariance matrix.
##              scale       shape
## scale  0.12612507 -0.03889909
## shape -0.03889909  0.01993603
##
```

```
##  AIC = 215.2792
##
##  BIC = 219.6585
```

```
ci(gpd, type="parameter")
```

```
## fevd(x = na.omit(dataWind$Speed), threshold = 20, type = "GP",
##     time.units = "months")
##
## [1] "Normal Approx."
##
##       95% lower CI   Estimate 95% upper CI
## scale    1.2005018  1.8965652     2.592629
## shape   -0.3161812 -0.0394441     0.237293
```

```
#GEV
ci(gev, type="return.level", return.period=20)
```

```
## fevd(x = max_monthly$Month)
##
## [1] "Normal Approx."
##
## [1] "20-year return level: 11.531"
##
## [1] "95% Confidence Interval: (9.4825, 13.5794)"
```

```
#GPD
ci(gpd, type="return.level", return.period=20)
```

```
## fevd(x = na.omit(dataWind$Speed), threshold = 20, type = "GP",
##     time.units = "months")
##
## [1] "Normal Approx."
##
## [1] "20-year return level: 20.704"
##
## [1] "95% Confidence Interval: (20.1867, 21.2218)"
```