

Hausaufgabe Wast3

J.Brändli, F.Kahlbacher, T.Haas

17 Dezember 2018

Inhaltsverzeichnis

1	Einleitung	1
2	Aufgaben	2
2.1	Aufgabe 1	2
2.2	Aufgabe 2	2
2.3	Aufgabe 3	2
2.4	Aufgabe 4	3
2.5	Aufgabe 5	5
2.6	Aufgabe 6	6
3	Fazit	6

1 Einleitung

2 Aufgaben

2.1 Aufgabe 1

2.2 Aufgabe 2

2.3 Aufgabe 3

2.4 Aufgabe 4

2.4 Aufgabe 4

In dieser Aufgabe ist die Variable Feinstaub (PM10) von Interesse, jeweils für die Messpunkte Stampfenbachstrasse, Schimmelstrasse und Rosengartenstrasse. Heubeeibüel hat keine Messungen des Feinstaubes und wird darum nicht berücksichtigt.

Der Tagesmittelgrenzwert vom PM10 beträgt $50\mu\text{g}/\text{m}^3$ und darf max. 1x pro Jahr ueberschritten werden.

Daten vorbereiten:

```
luftqual.PM10 <- luftqual %>% ungroup() %>%  
  select(Datum, 'Feinstaub PM10', Station) %>% # nur relevante Kolumnen  
  filter(Station != "Heubeeibüel") %>%  
  mutate(PM10_uberschritt = `Feinstaub PM10` >= 50 )  
  
head(luftqual.PM10)
```

```
## # A tibble: 6 x 4  
##   Datum      `Feinstaub PM10` Station    PM10_uberschritt  
##   <date>          <dbl> <chr>      <lgl>  
## 1 2012-01-01      NaN Rosengarten NA  
## 2 2012-01-02      NaN Rosengarten NA  
## 3 2012-01-03      NaN Rosengarten NA  
## 4 2012-01-04      NaN Rosengarten NA  
## 5 2012-01-05      NaN Rosengarten NA  
## 6 2012-01-06      NaN Rosengarten NA
```

Die ersten 6 Werte sind per Zufall gerade NA.

2.4.1 Wie oft wird der Tagesmittel-Grenzwert an welcher Station überschritten?

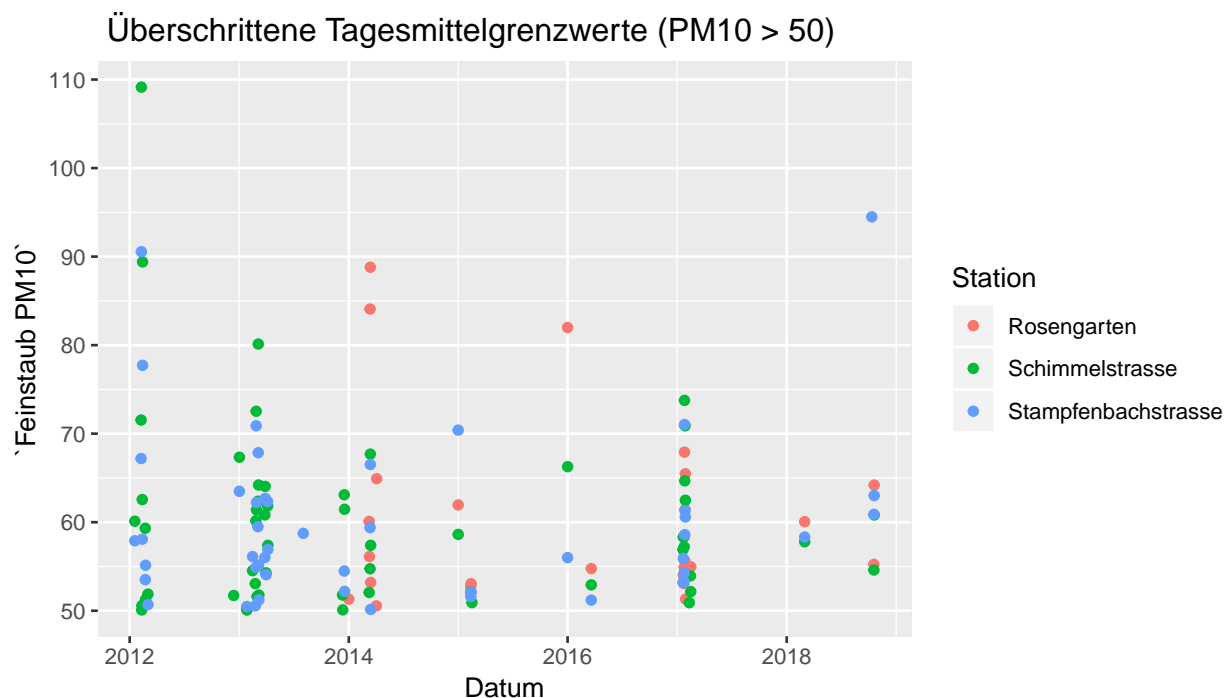
```
luftqual.PM10 %>% group_by(Station) %>% summarize(n = sum(PM10_uberschritt, na.rm = T))  
  
## # A tibble: 3 x 2  
##   Station      n  
##   <chr>    <int>  
## 1 Rosengarten    25  
## 2 Schimmelstrasse 56  
## 3 Stampfenbachstrasse 47
```

An der Schimmelstrasse wird der Tagesmittel-Grenzwert am häufigsten überschritten.

2.4 Aufgabe 4

2.4.2 Überschrittenen Tagesmittelgrenzwerte im Zeitlichen verlauf für alle Stationen in einer Grafik

```
luftqual.PM10 %>% filter(PM10_überschritt) %>%  
  ggplot(aes(x=Datum, y=`Feinstaub PM10`)) + geom_point(aes(color = Station)) + ggtitle("Überschrittenen Tagesmittelgrenzwerte (PM10 > 50)")
```



Der Tagesmittelgrenzwert wird bei allen Stationen oft in den gleichen Wochen / Monaten uebertroffen. Dies ist gut moeglich, da alle Messstationen ähnlichen Einflüssen ausgesetzt sind, auf welche der Feinstaub reagiert. Wie z.B. die Jahreszeit und das Wetter.

Aus der Grafik ist ersichtlich, dass die Feinstaubbelastung im Winter höher ist, als in den anderen Jahreszeiten.

2.4.3 In welchen Jahren und Stationen ist der Anteil der Tage mit Grenzwert überschreitungen signifikant grösser als zufällig

Jahresmittelgrenzwert = $20 \mu\text{g}/\text{m}^3$

Daten vorbereiten: Nach dem Jahr und Station gruppieren

```
luftqual.PM10 <- luftqual.PM10 %>% mutate(Jahr = strtrim(luftqual.PM10$Datum, 4)) %>%  
  group_by(Jahr, Station) %>%  
  summarize(n = sum(PM10_überschritt, na.rm = T))  
  
head(luftqual.PM10)
```

```
## # A tibble: 6 x 3  
## # Groups:   Jahr [2]  
##   Jahr Station n
```

2.5 Aufgabe 5

```
##   <chr> <chr>           <int>
## 1 2012  Rosengarten      0
## 2 2012  Schimmelstrasse 11
## 3 2012  Stampfenbachstrasse 8
## 4 2013  Rosengarten      0
## 5 2013  Schimmelstrasse 22
## 6 2013  Stampfenbachstrasse 19
```

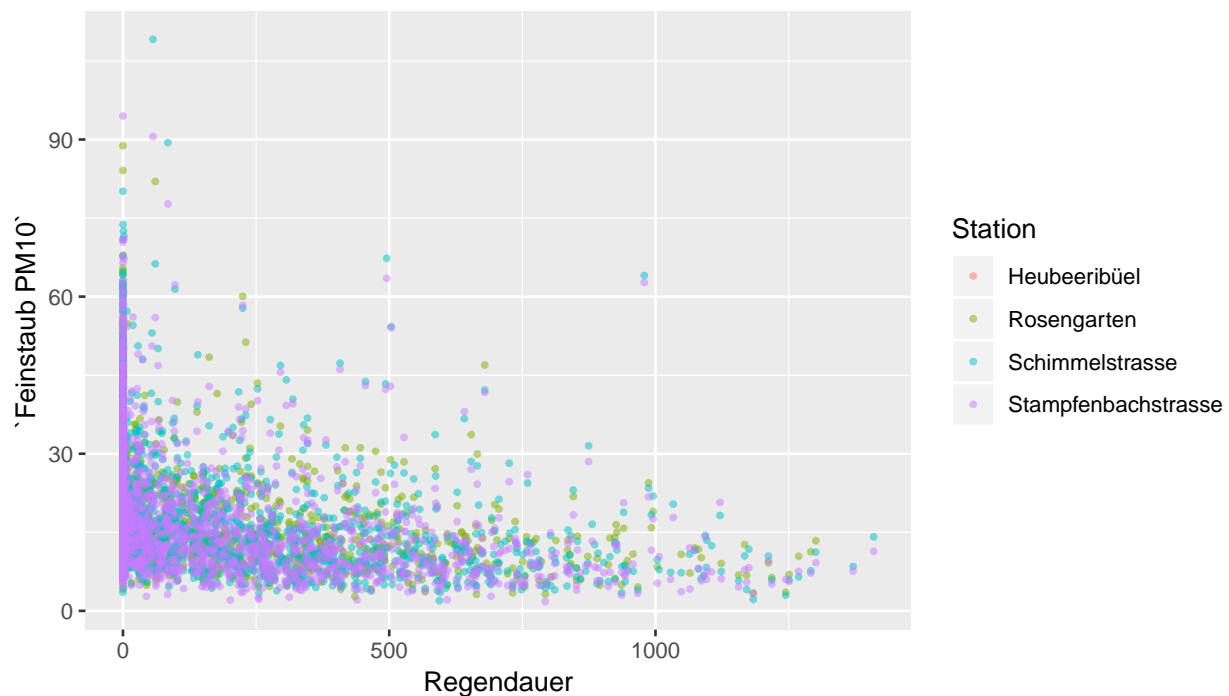
```
# pairwise wilcox test
# pairwise.wilcox.test(luftqual.PM10$n, luftqual.PM10$Jahr, luftqual.PM10$Station)
```

2.5 Aufgabe 5

Visualisierung des Zusammenhanges zwischen der Regenmenge und der Feinstaubkonzentration pro Station.

```
ordered_PM10 <- select(luftqual, Datum, `Feinstaub PM10`, Station, Regendauer)

# Plot
ggplot(ordered_PM10, aes(x= Regendauer, y= `Feinstaub PM10`)) +
  geom_point(aes(color = Station), alpha = 0.5, size = 0.8)
```



Auf der Grafik ist folgendes Verhalten zu erkennen: Die Feinstaubkonzentration ist tendenziell höher wenn es nicht Regnet, als wenn es Regnet. Um so länger es Regnet, nimmt dieser Effekt noch zu.

2.6 Aufgabe 6

Mit einem Statistischen Test soll überprüft werden, ob ein Zusammenhang besteht. Es wurde der t-Test verwendet, da die Mittelwerte der Feinstaubkonzentration normalverteilt aber die Varianz nicht bekannt sind.

h_0 : $\mu_{\text{Feinstaub_regen}} = \mu_{\text{Feinstaub_keinregen}}$

h_1 : $\mu_{\text{Feinstaub_regen}} < \mu_{\text{Feinstaub_keinregen}}$

```
PM10_test <- ordered_PM10 %>% group_by(Datum) %>%
  summarise(PM10_mean = mean(`Feinstaub PM10`, na.rm = T) ,
    Regendauer = Regendauer[1])
```

```
# tTest
```

```
t.test(x = PM10_test$PM10_mean[PM10_test$Regendauer != 0],
  y = PM10_test$PM10_mean[PM10_test$Regendauer == 0],
  conf.level = 0.99, alternative = "less")
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: PM10_test$PM10_mean[PM10_test$Regendauer != 0] and PM10_test$PM10_mean[PM10_test$Regendauer == 0]
```

```
## t = -16.808, df = 2027.1, p-value < 2.2e-16
```

```
## alternative hypothesis: true difference in means is less than 0
```

```
## 99 percent confidence interval:
```

```
##      -Inf -6.078394
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 16.08592 23.14162
```

$p < 0.01 \rightarrow h_0$ verwerfen

Die Nullhypothese wird klar verworfen. Das heisst, die Feinstaubkonzentration ist an Tagen mit Regen signifikant tiefer als an Tagen mit Regen.

2.6 Aufgabe 6

3 Fazit