World Marathon Majors

Erste Analysen

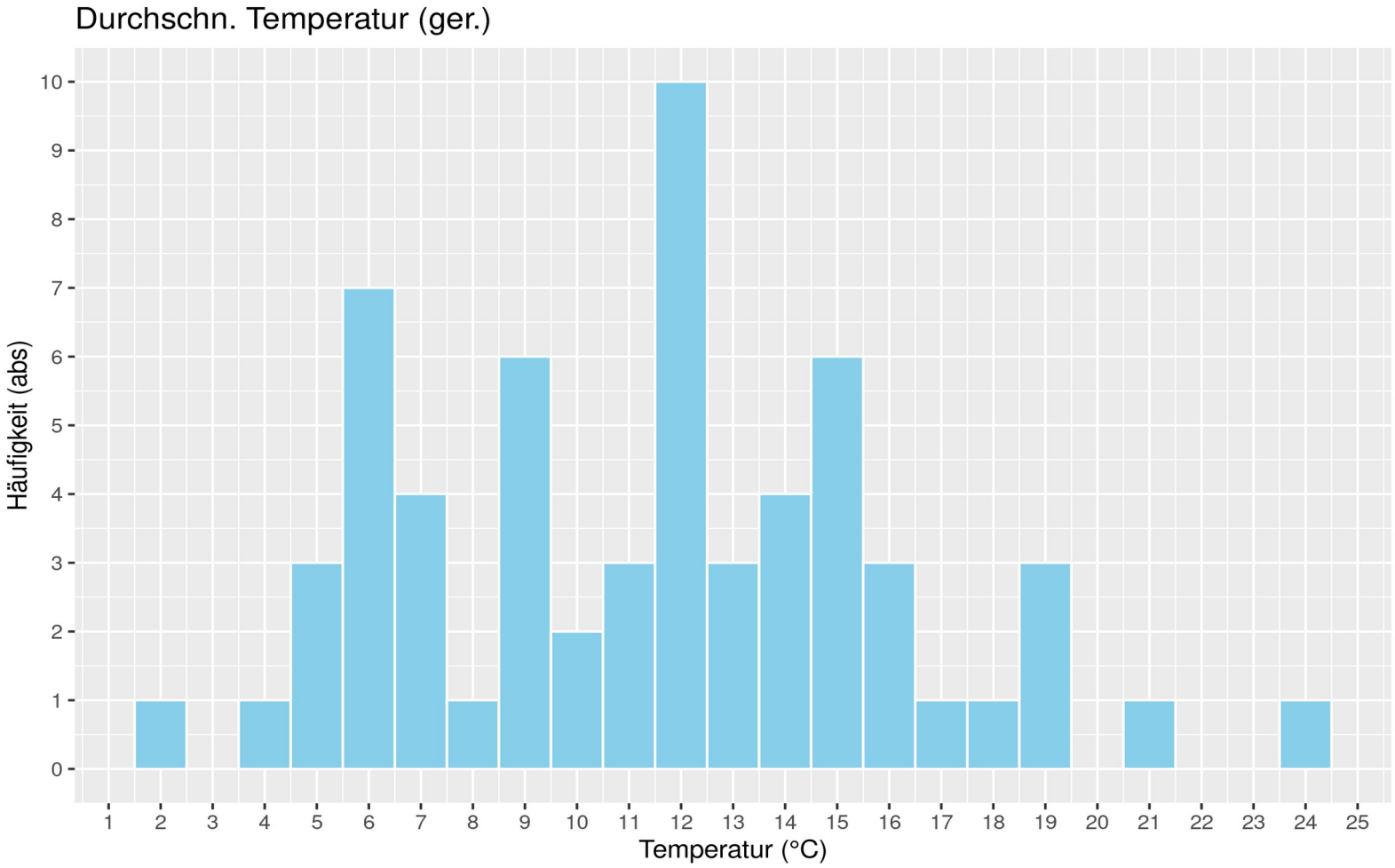
Paul Heiser

November 2021

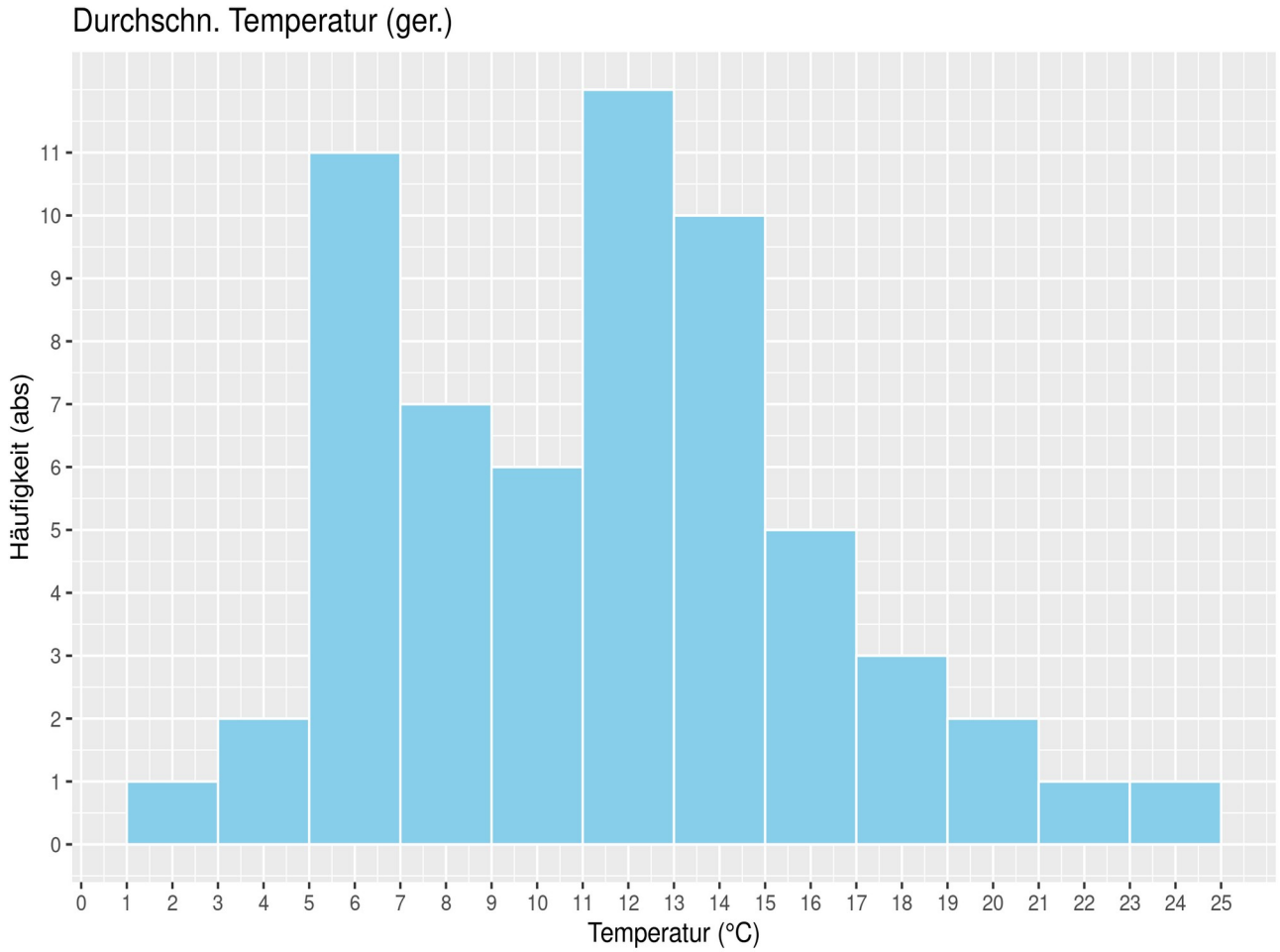
Wetterdaten – R-Summary

	Windstärke		Temperatur		Taupunkt	
	MEAN	MEAN_RND	MEAN	MEAN_RND	MEAN	MEAN_RND
Min.	0.375	0.40	1.375	1.50	-11.162	-11.000
Q1	2.742	2.70	7.417	7.50	-0.500	-0.500
Median	3.600	3.60	12.167	12.00	4.535	4.500
Mean	3.926	3.92	11.656	11.64	4.413	4.402
Q3	4.900	4.90	14.900	15.00	9.833	10.000
Max.	10.450	10.50	24.575	24.50	19.000	19.000

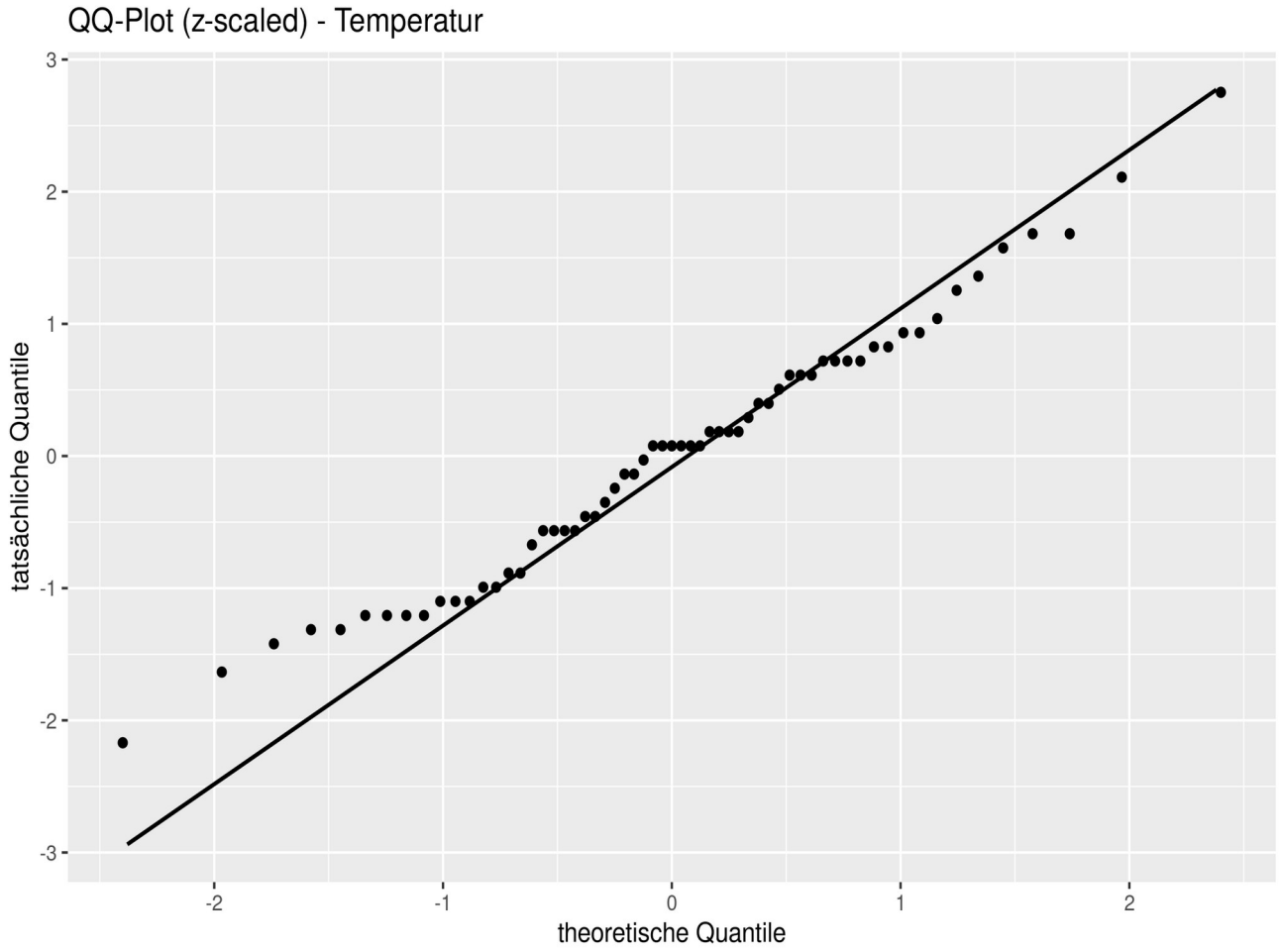
Wetterdaten – Temperatur (1)



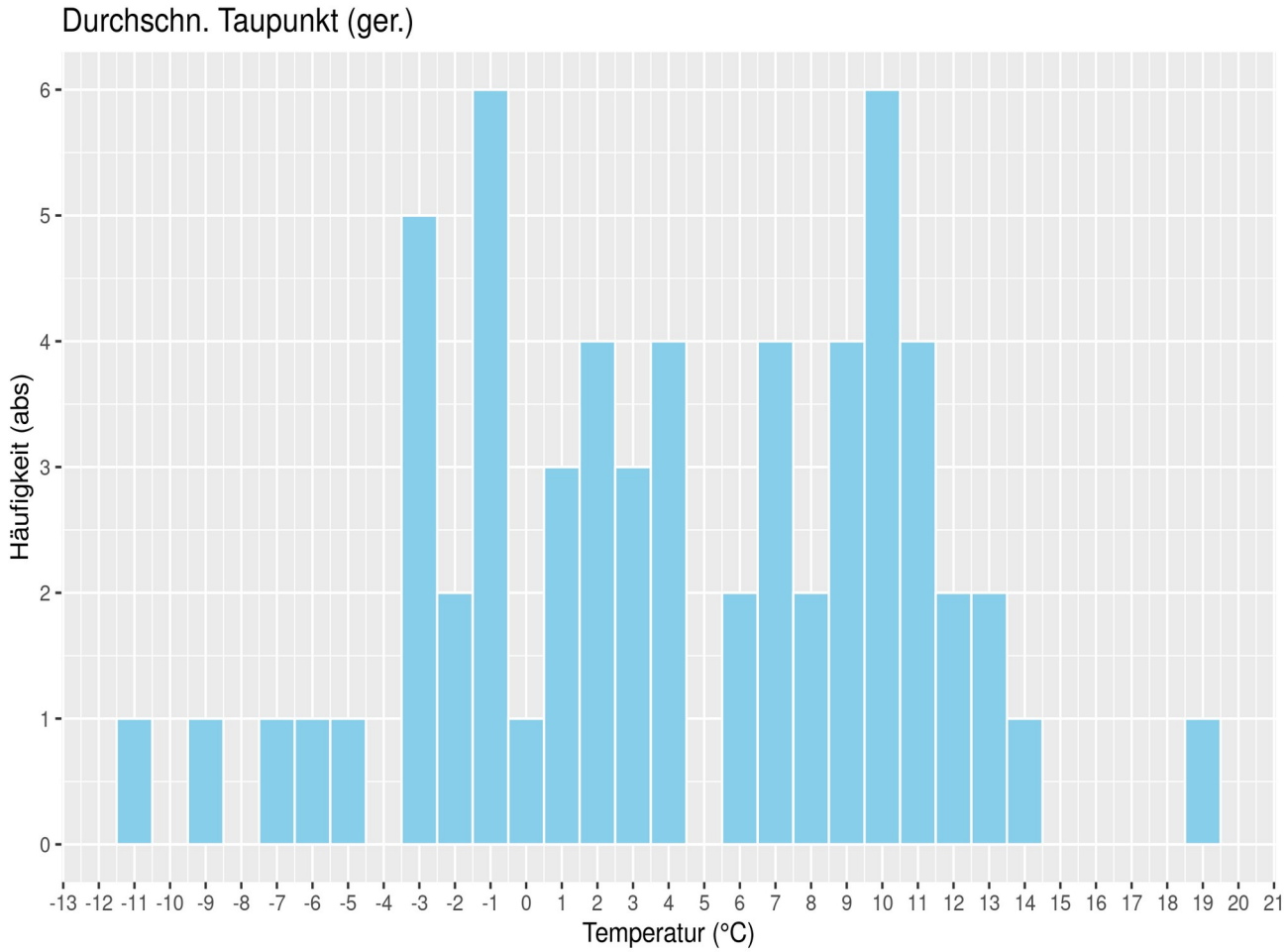
Wetterdaten – Temperatur (2)



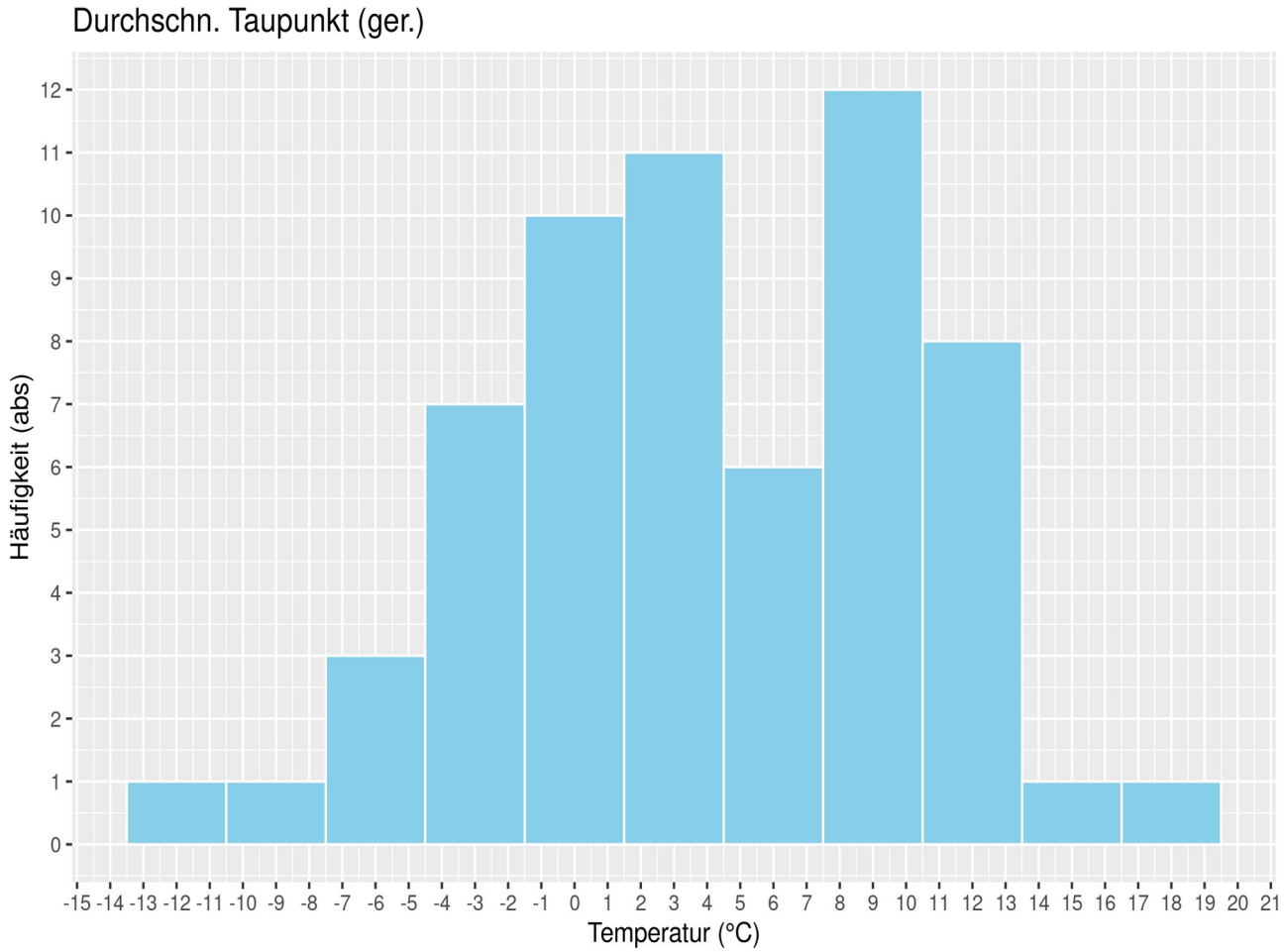
Wetterdaten – Temperatur (3)



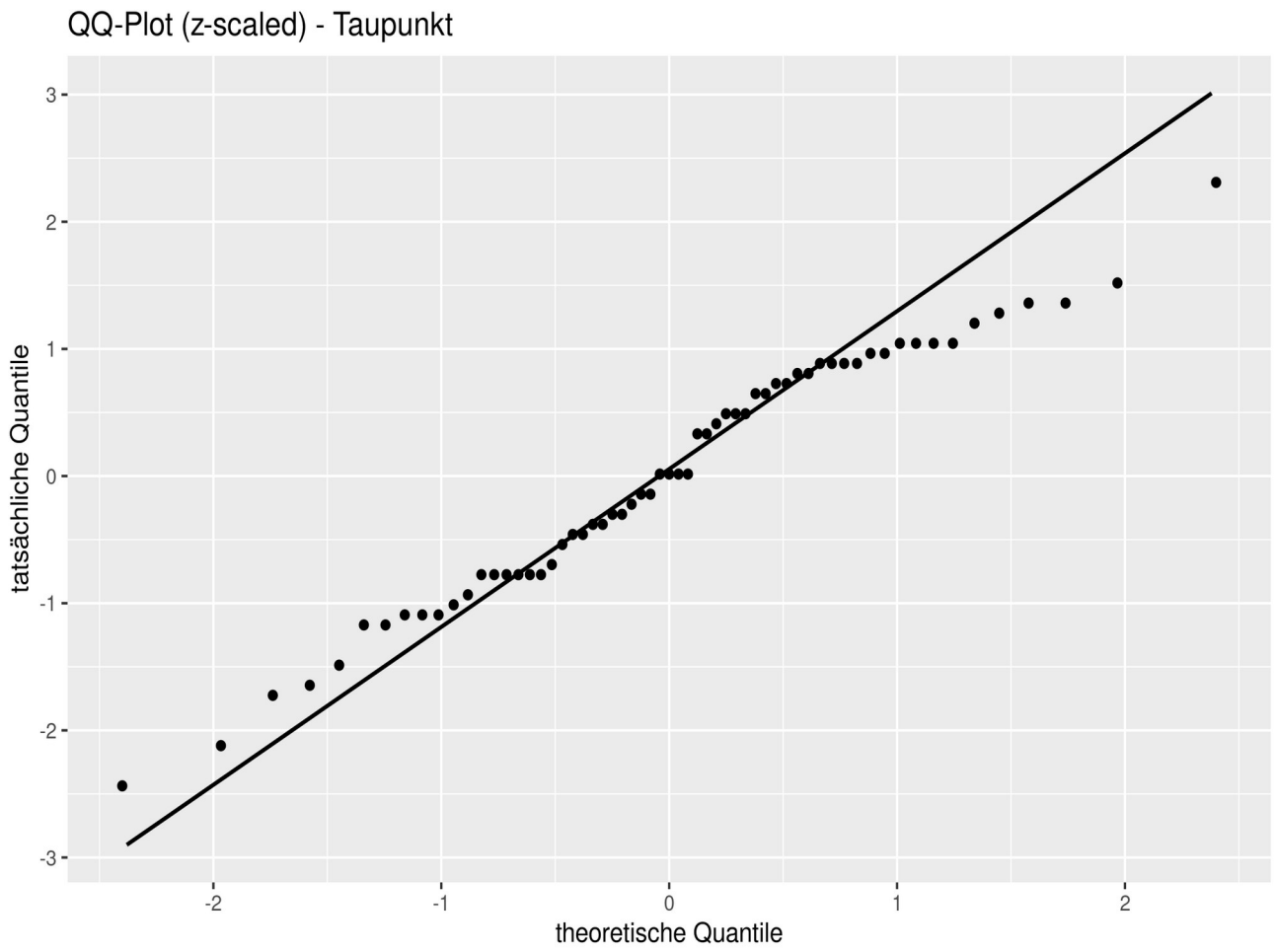
Wetterdaten – Taupunkt (1)



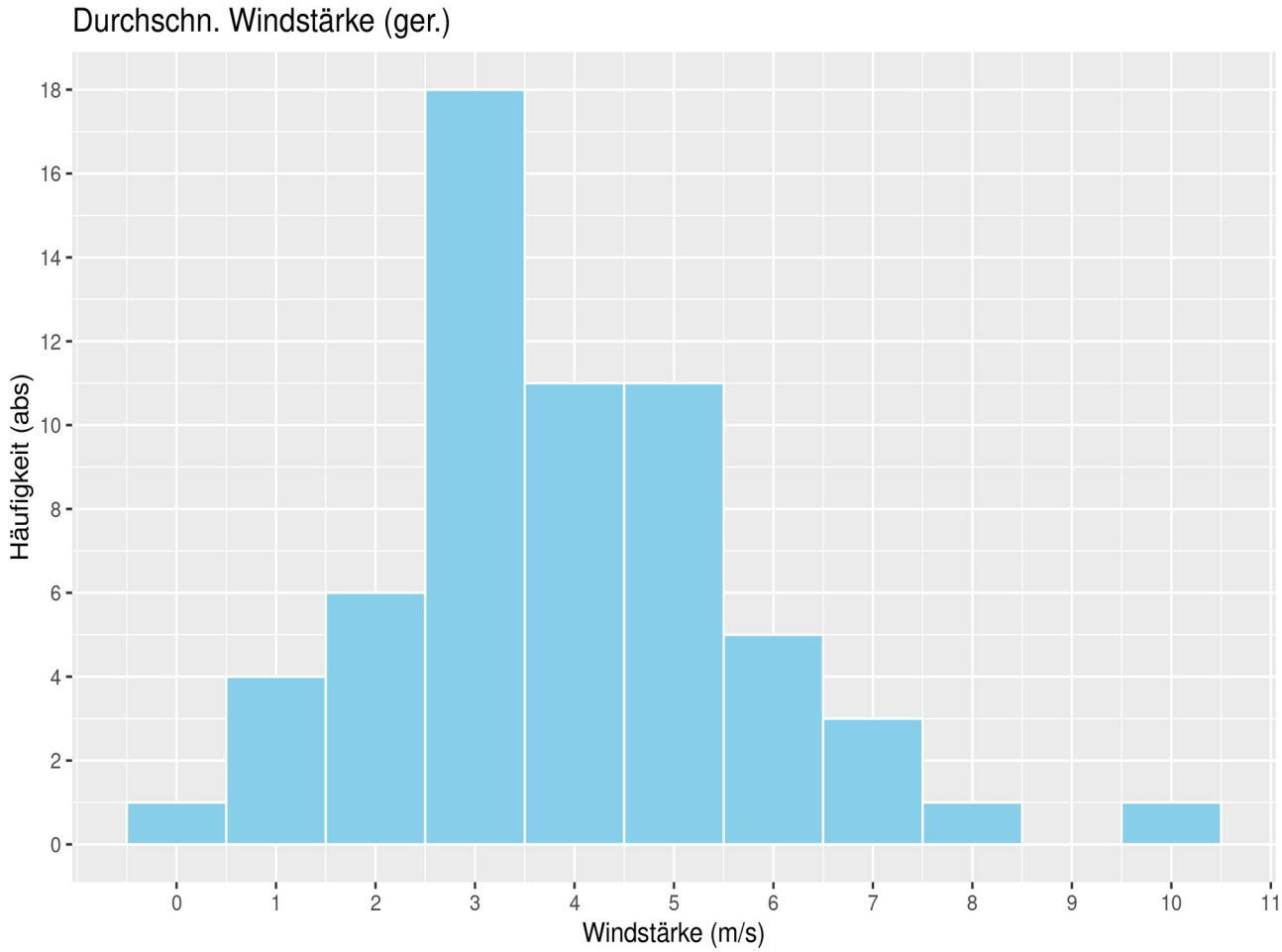
Wetterdaten – Taupunkt (2)



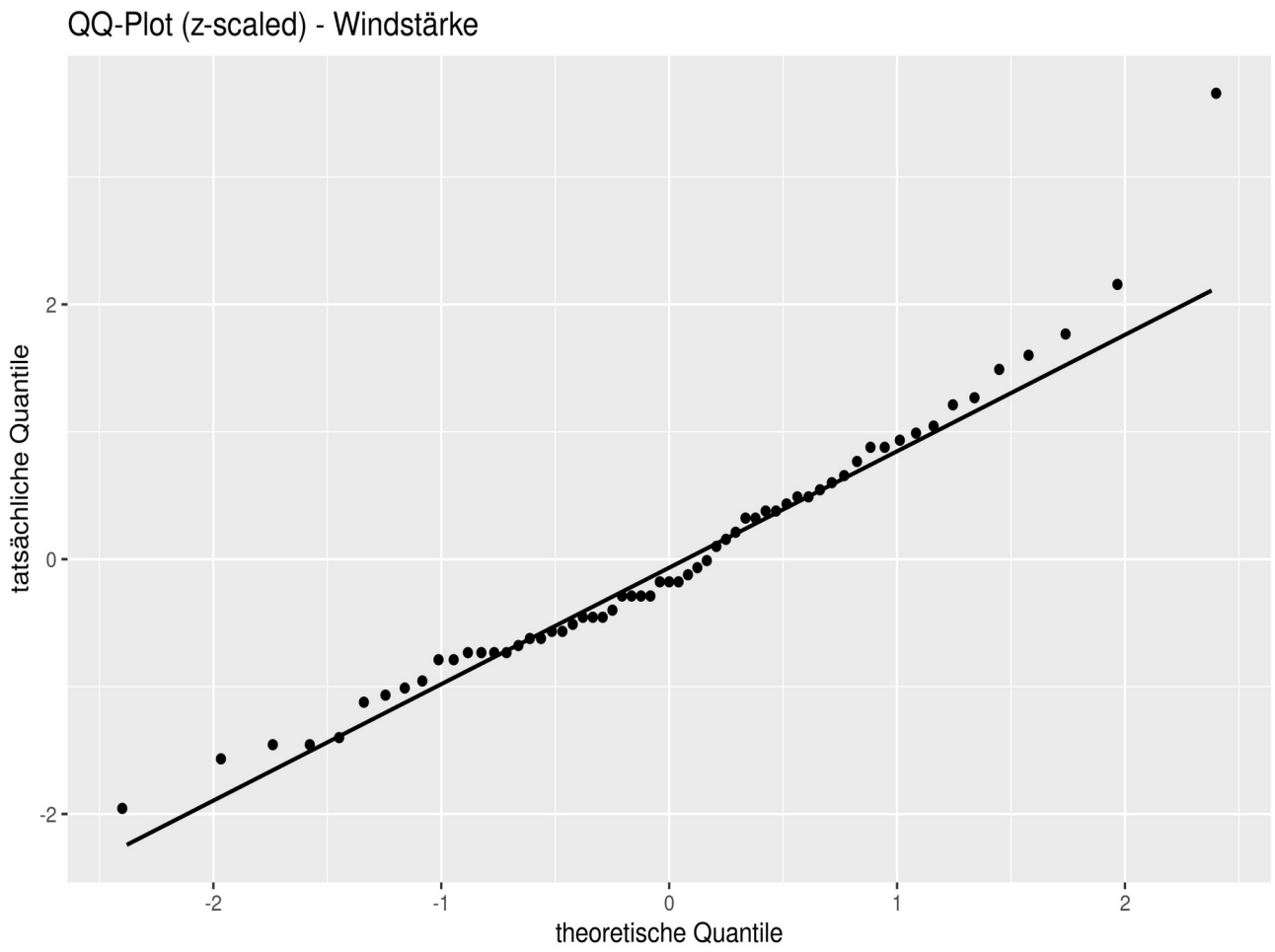
Wetterdaten – Taupunkt (3)



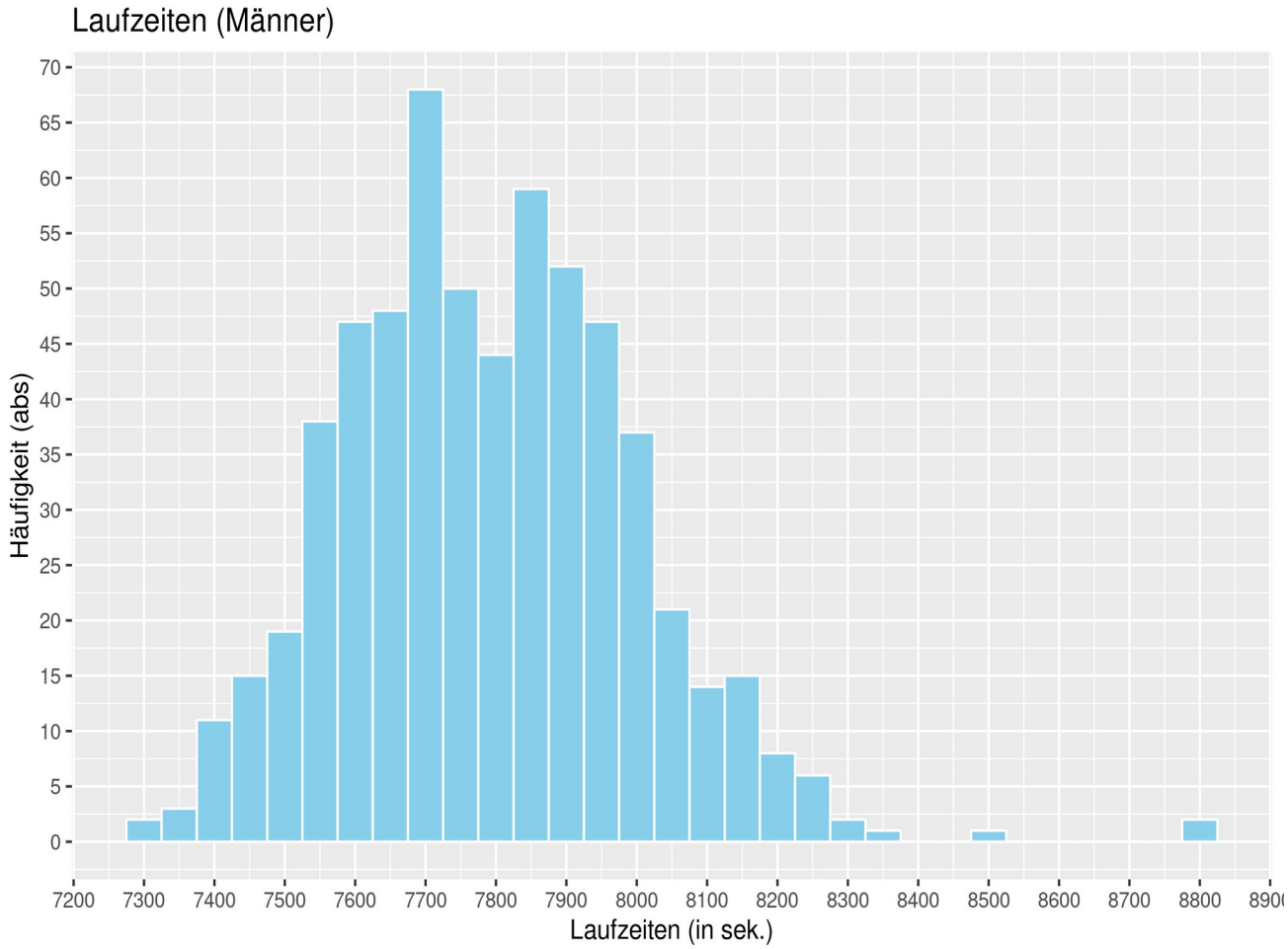
Wetterdaten – Windstärke (1)



Wetterdaten – Windstärke (2)

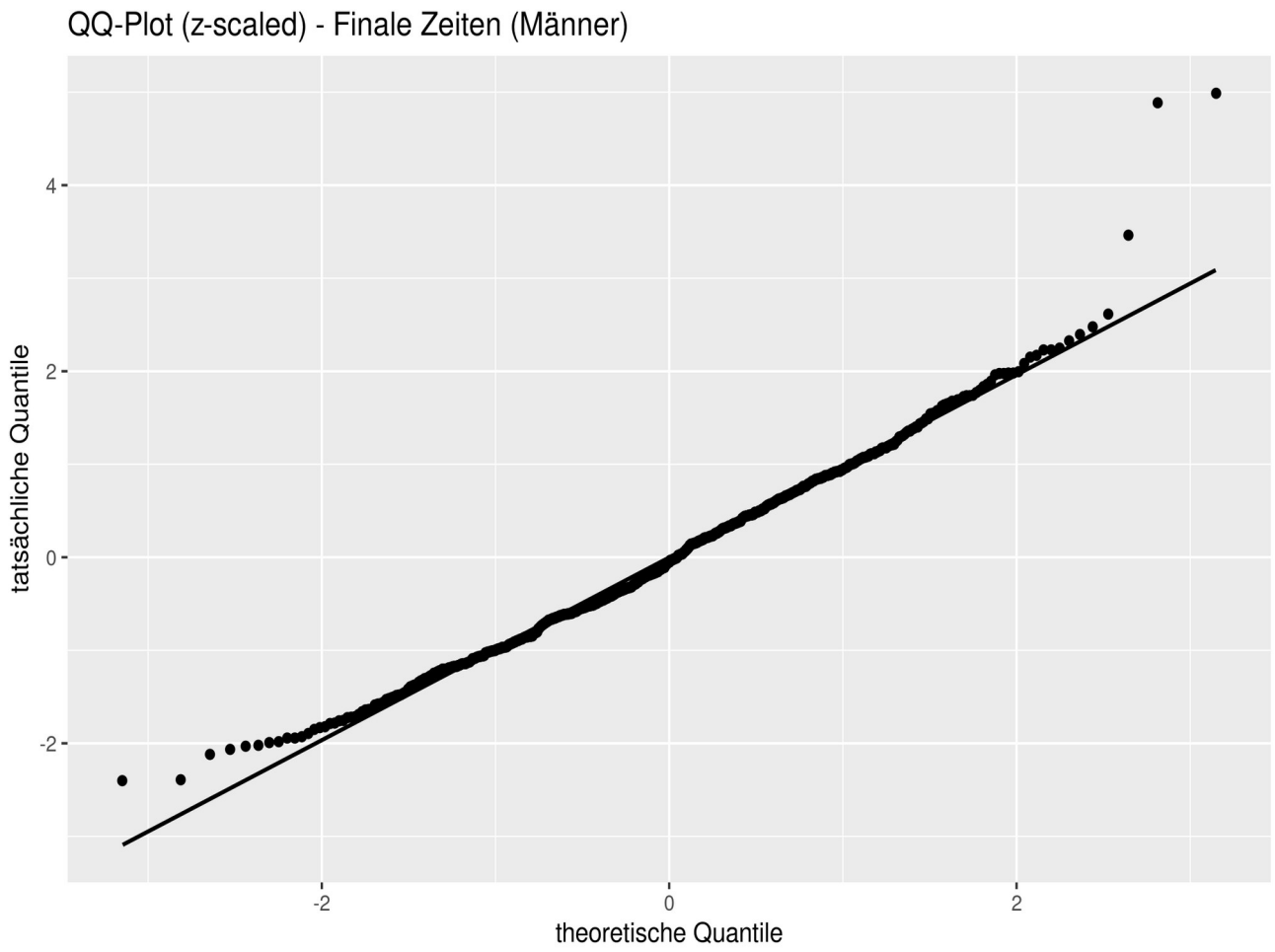


Marathondaten – Verteilung der Laufzeiten (Männer) (1)

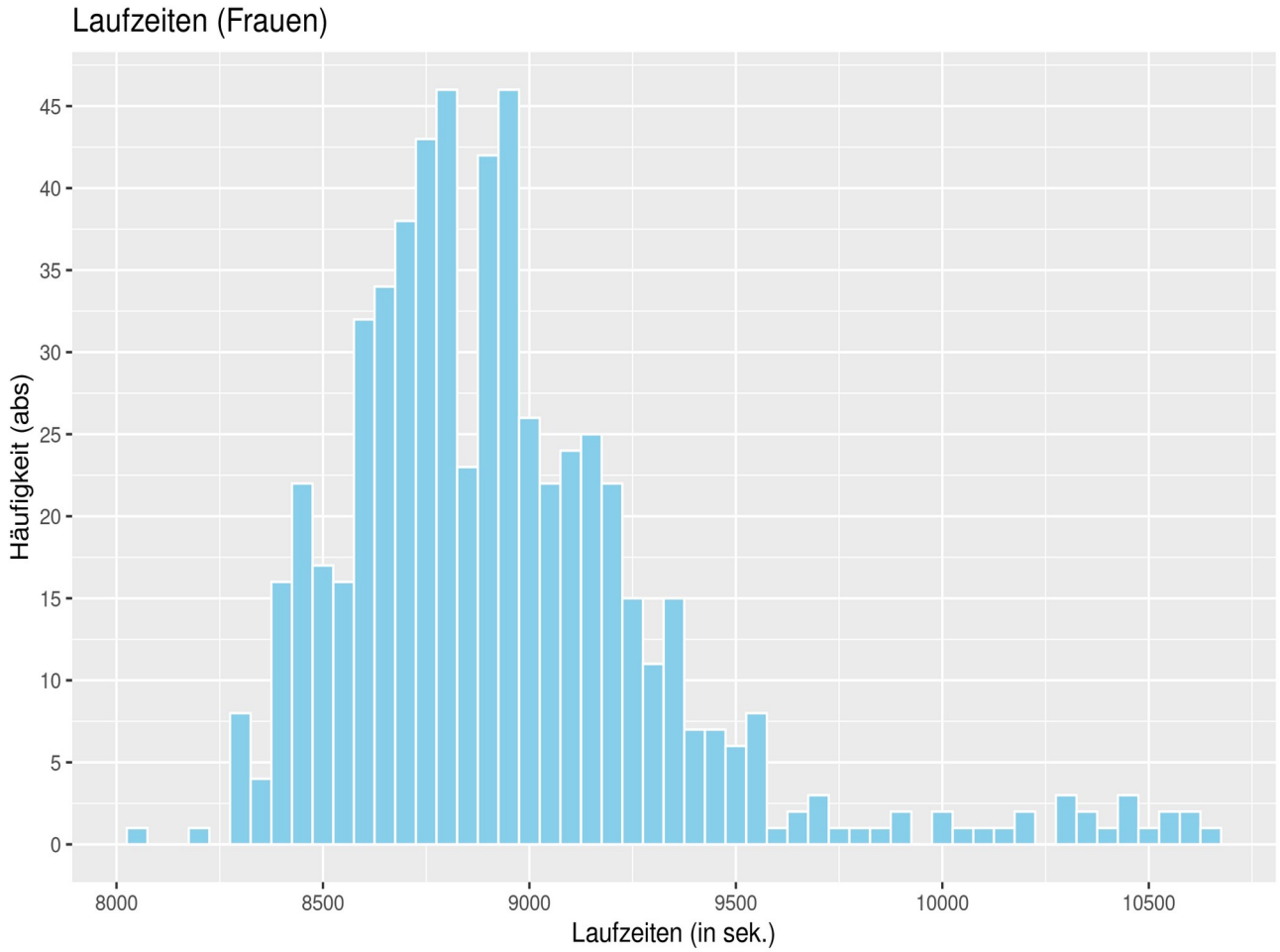


Min.:	7299
1Q:	7656
Median:	7781
Mean:	7792
3Q:	7927
Max.:	8815

Marathondaten – Verteilung der Laufzeiten (Männer) (2)

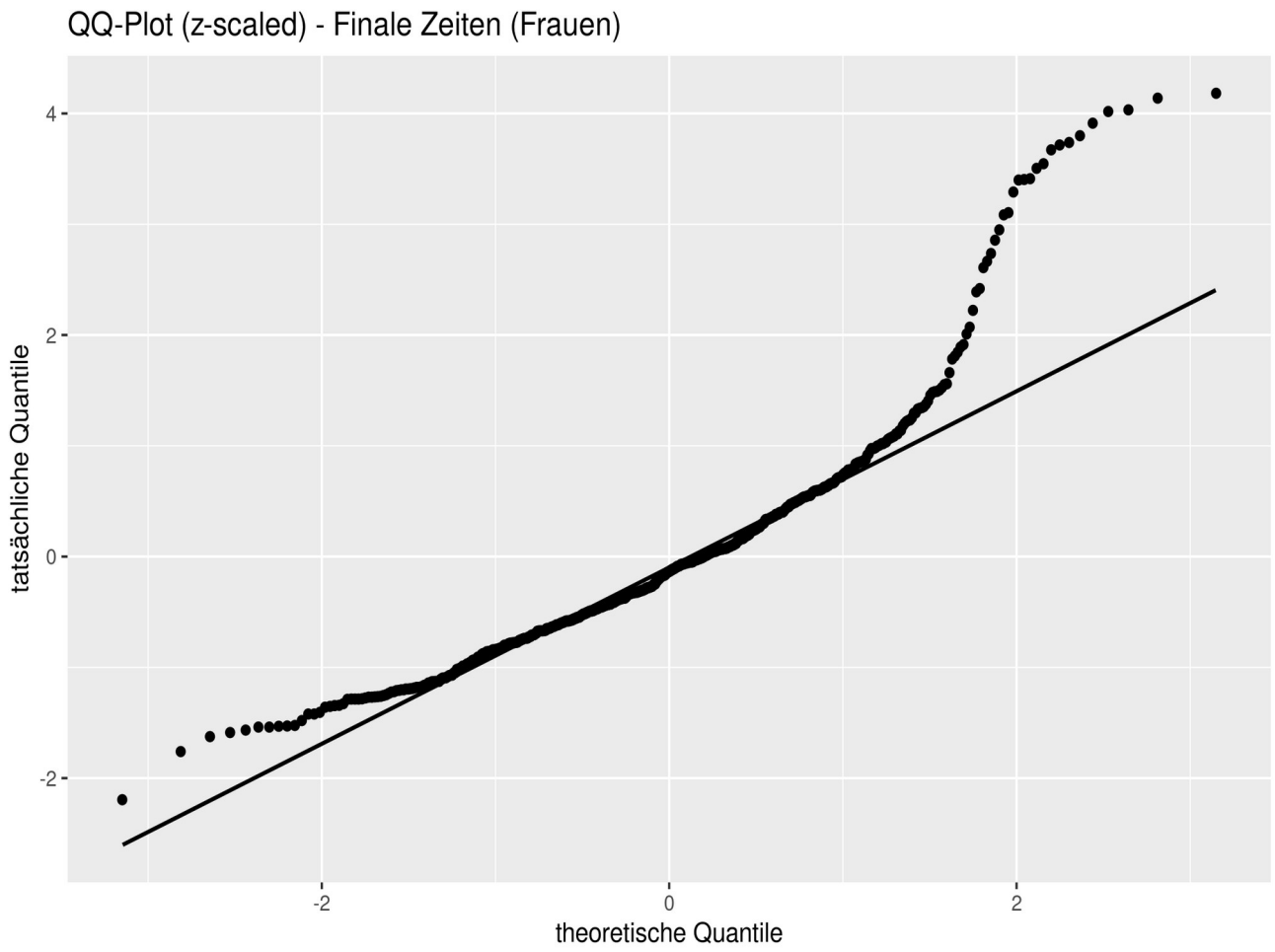


Marathondaten – Verteilung der Laufzeiten (Frauen) (1)

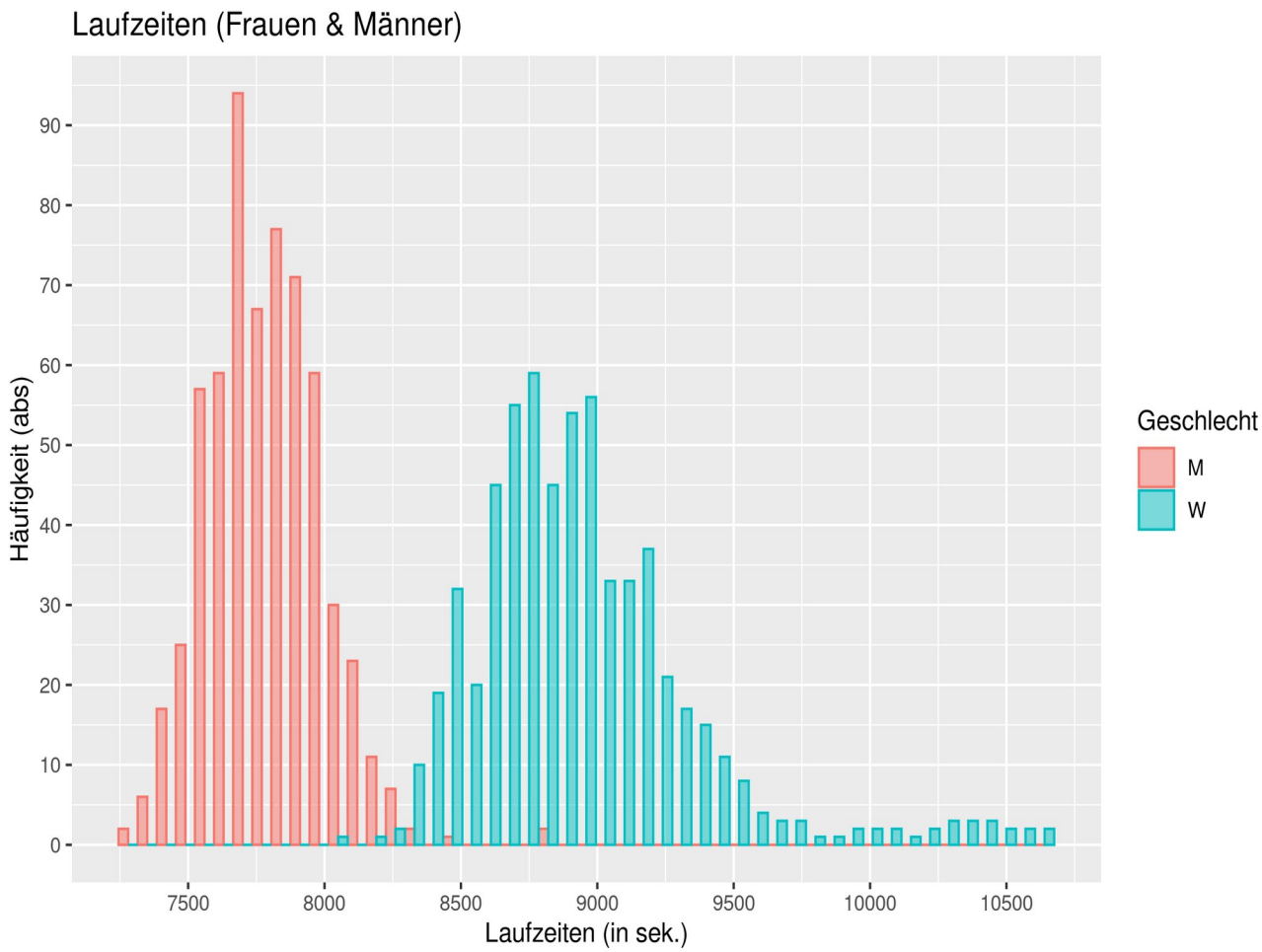


Min.: 8044
1Q: 8679
Median: 8883
Mean: 8938
3Q: 9116
Max.: 10641

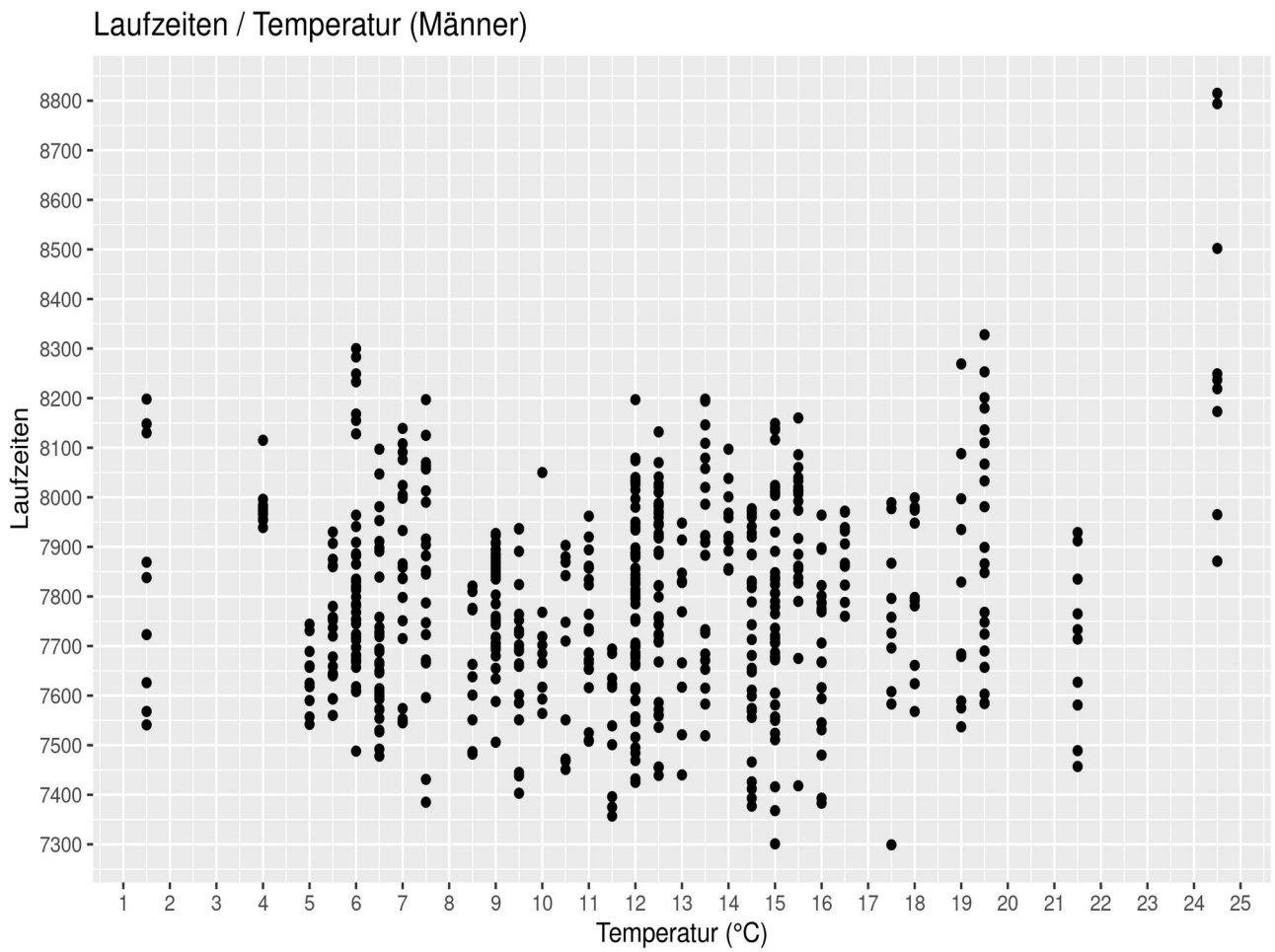
Marathondaten – Verteilung der Laufzeiten (Frauen) (2)



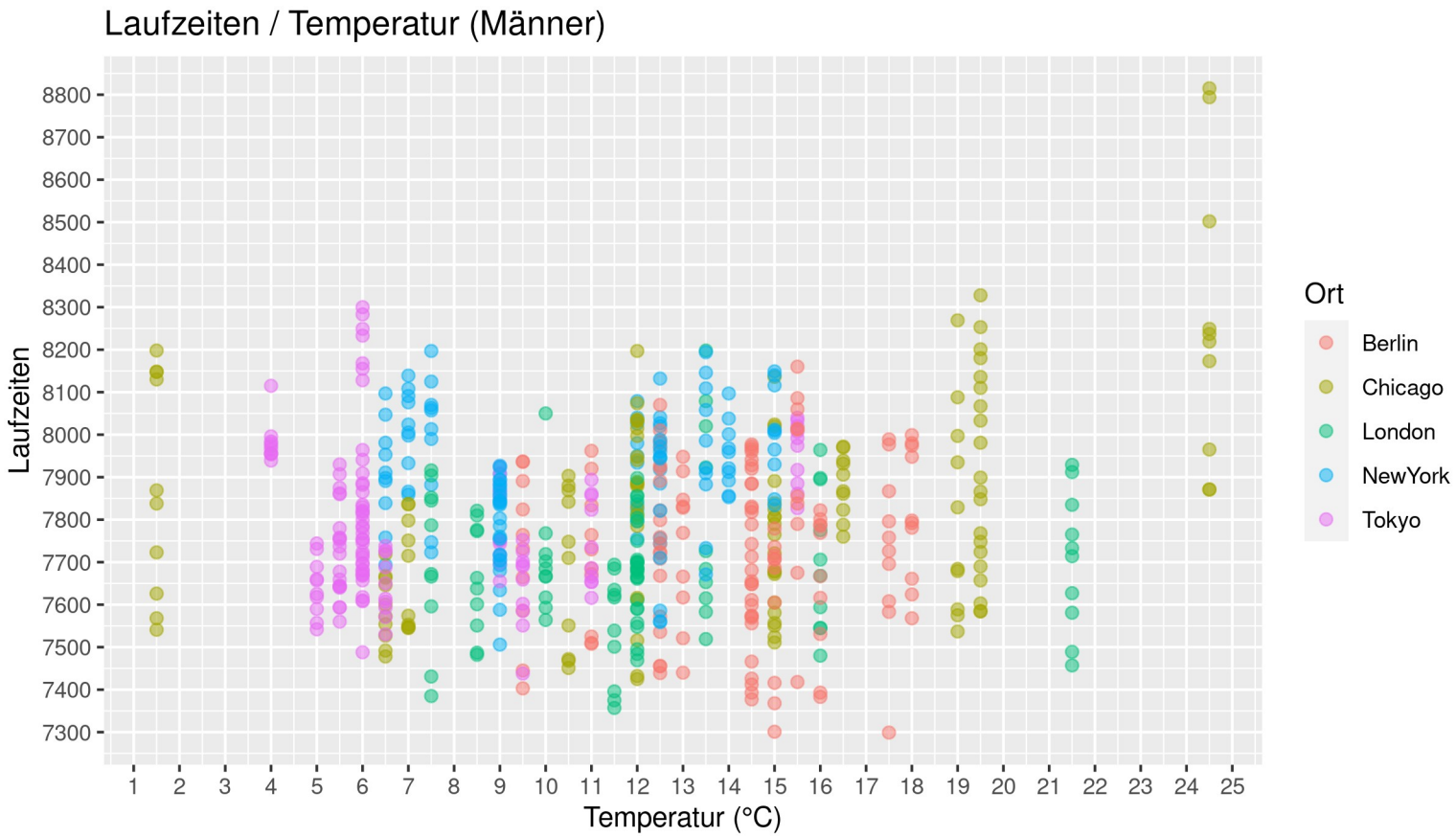
Marathondaten – Verteilung der Laufzeiten (Frauen & Männer)



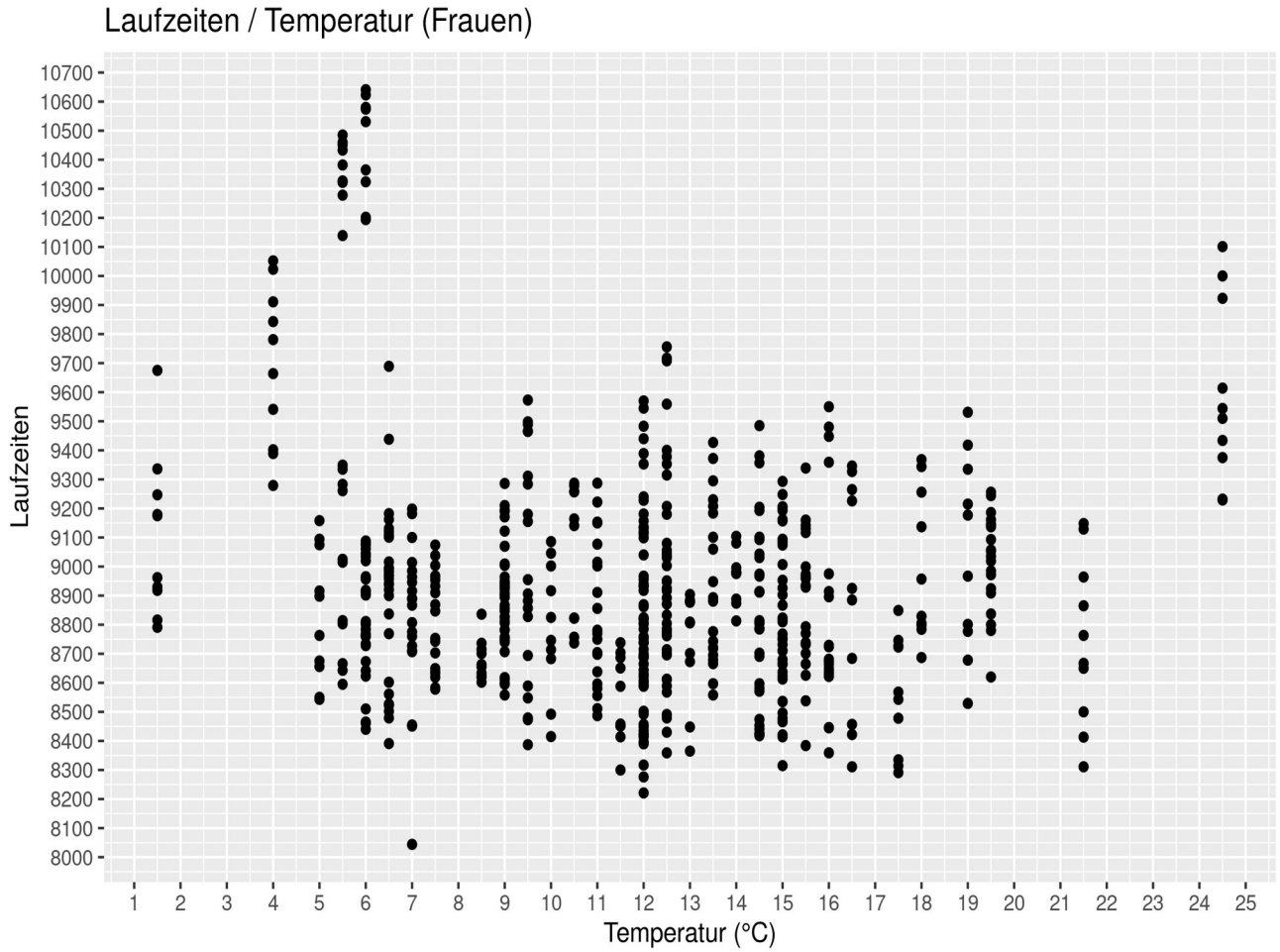
Marathondaten – Temperatur / Laufzeiten (Männer) (1)



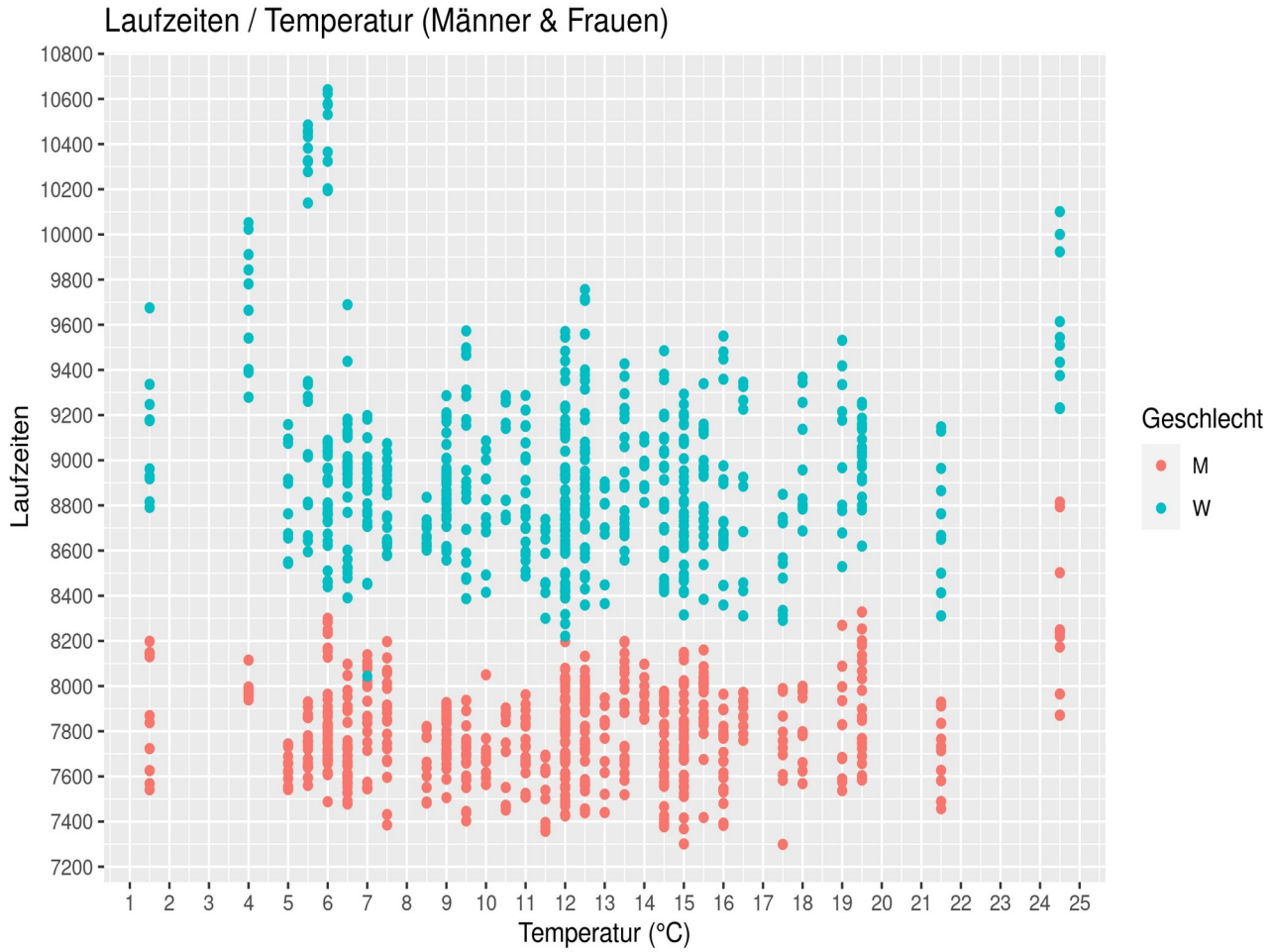
Marathondaten – Temperatur / Laufzeiten (Männer) (2)



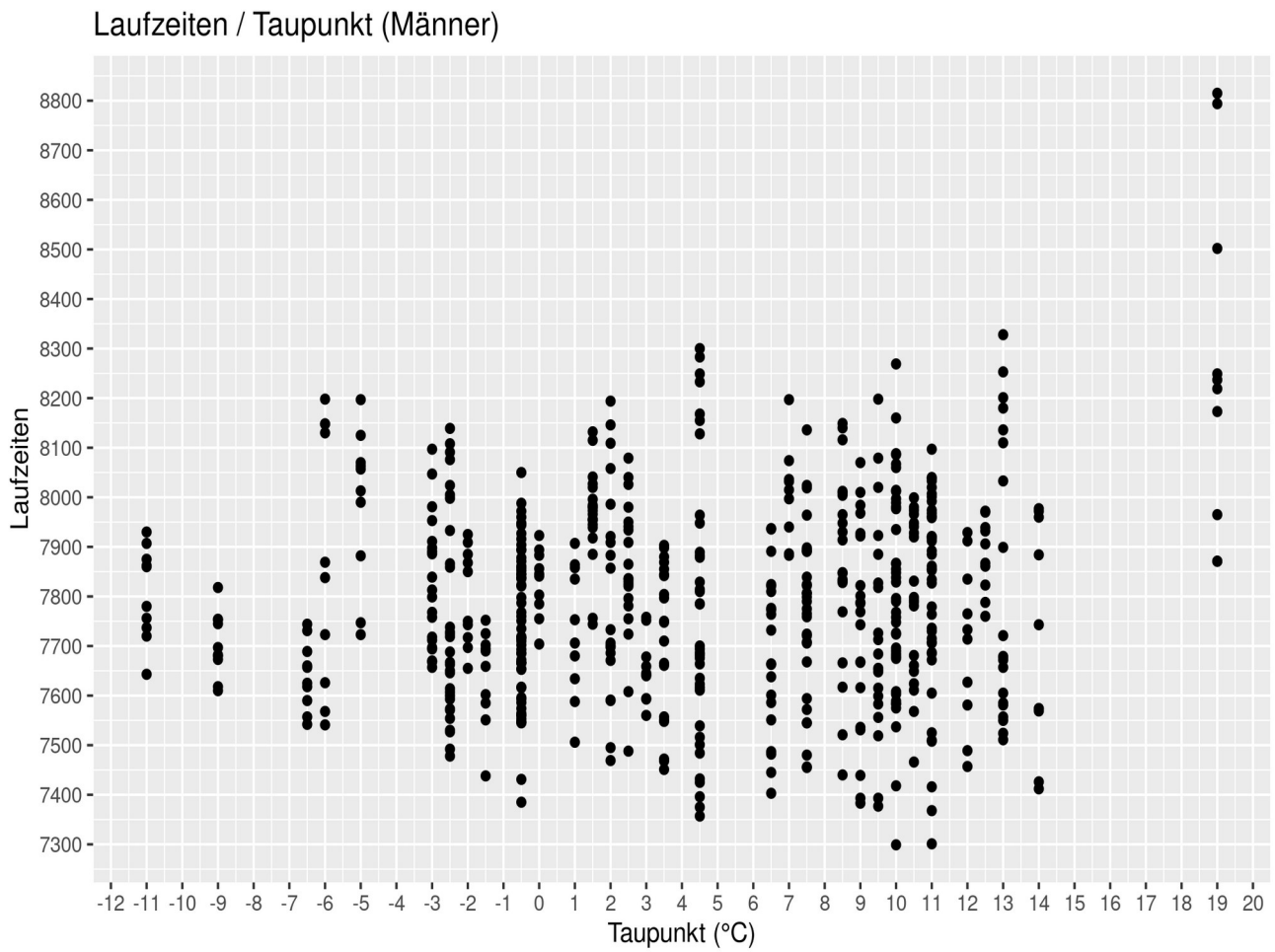
Marathondaten – Temperatur / Laufzeiten (Frauen)



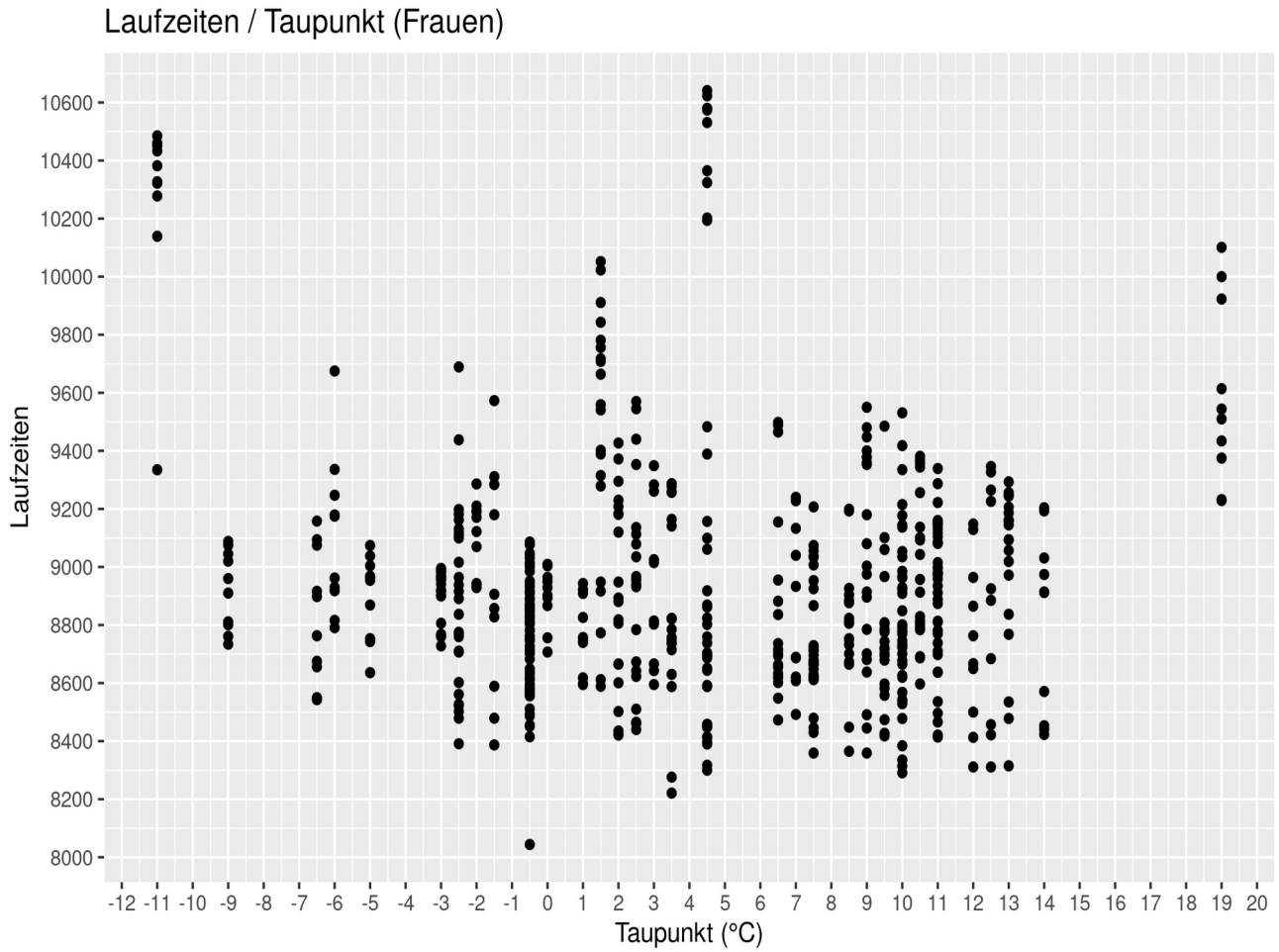
Marathondaten – Temperatur / Laufzeiten (Frauen & Männer)



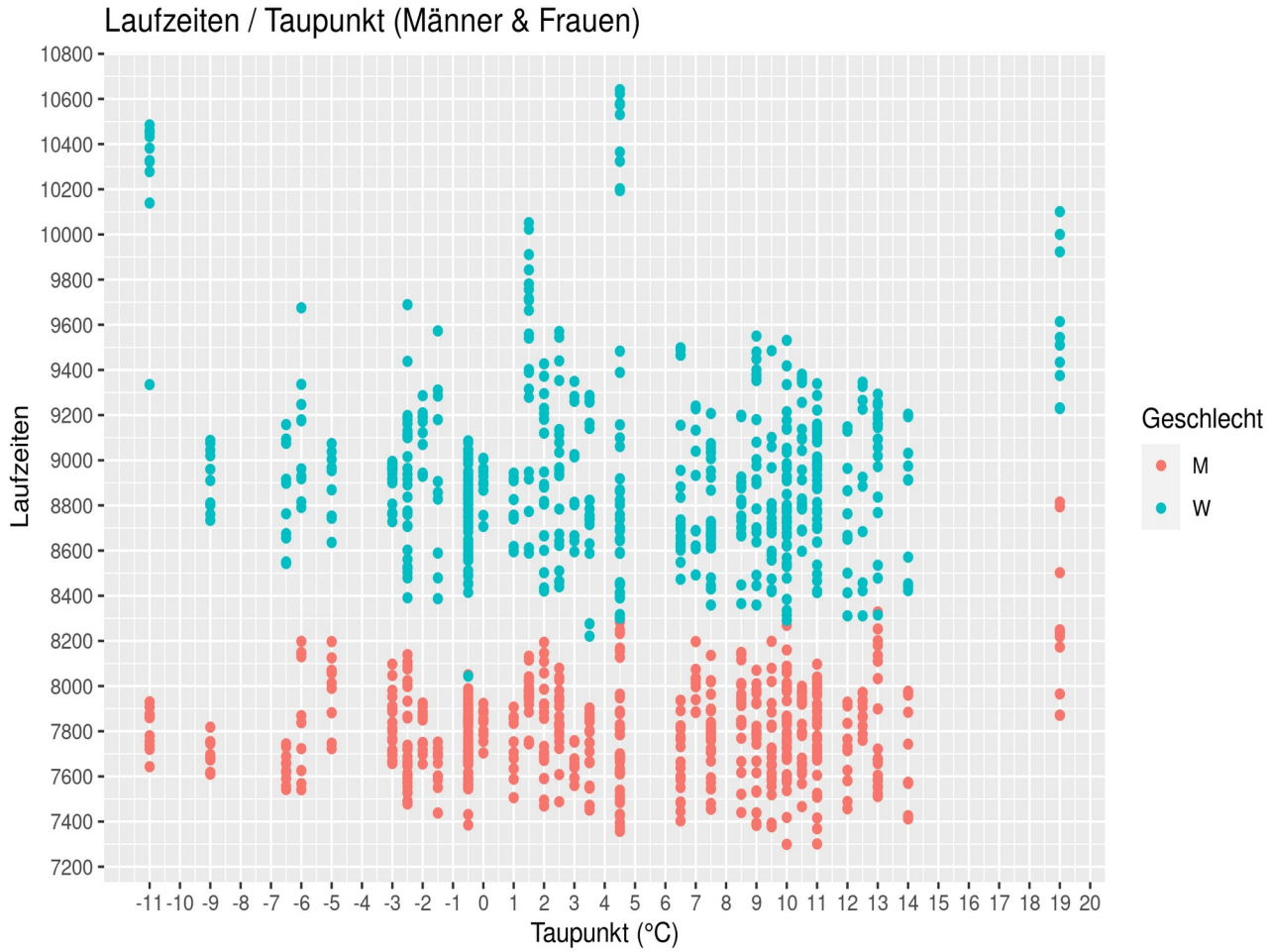
Marathondaten – Taupunkt / Laufzeiten (Männer)



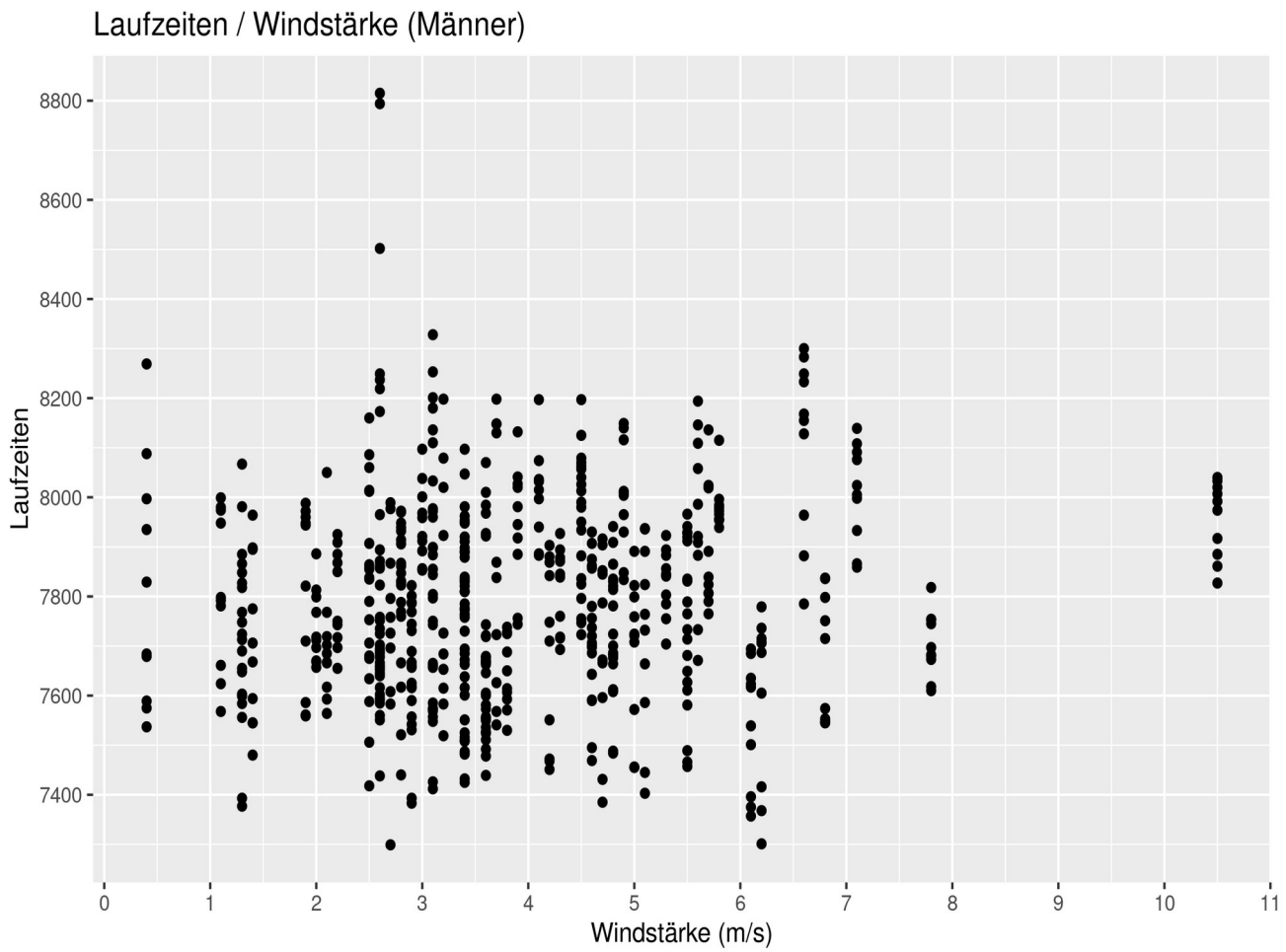
Marathondaten – Taupunkt / Laufzeiten (Frauen)



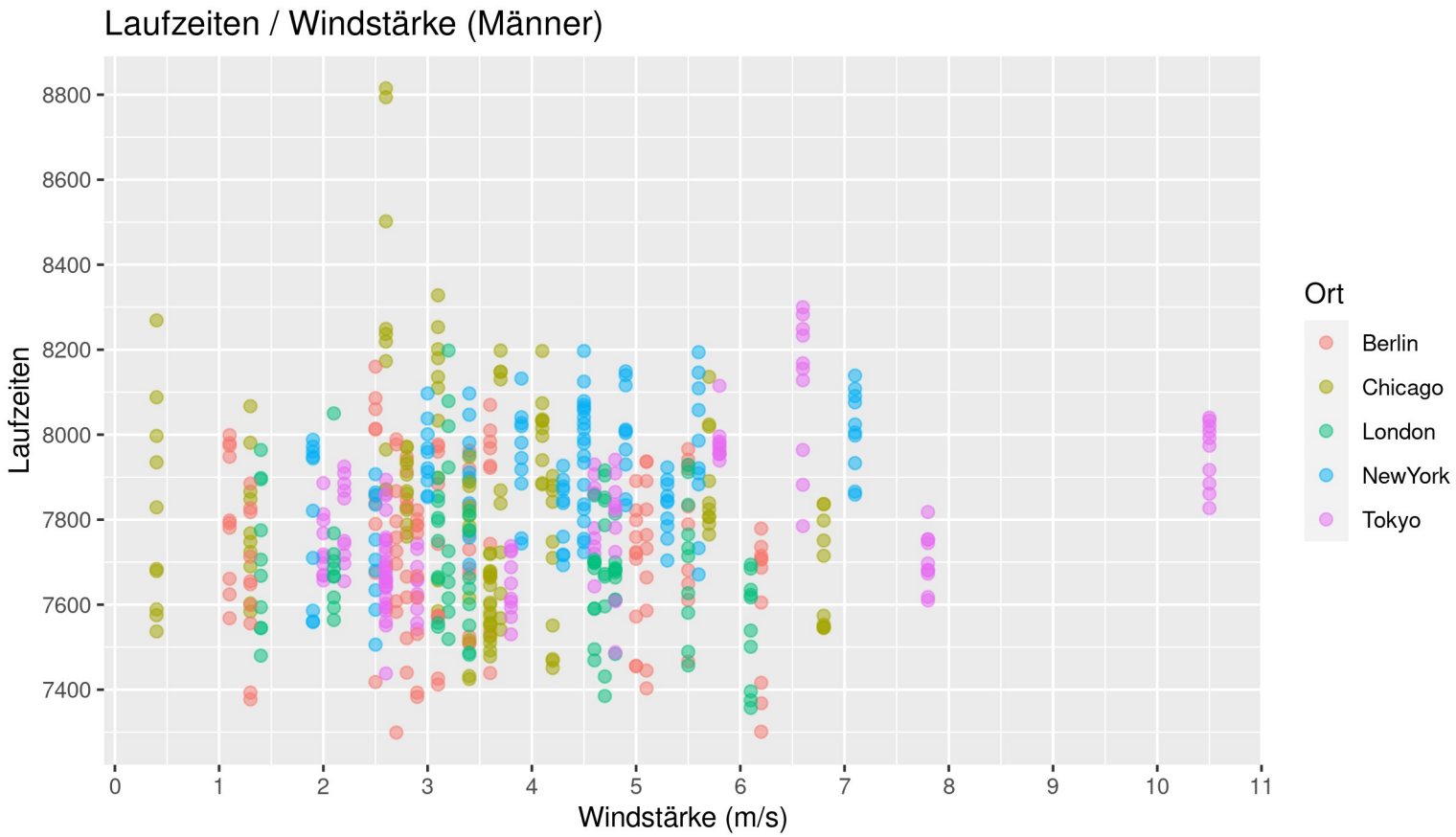
Marathondaten – Taupunkt / Laufzeiten (Frauen & Männer)



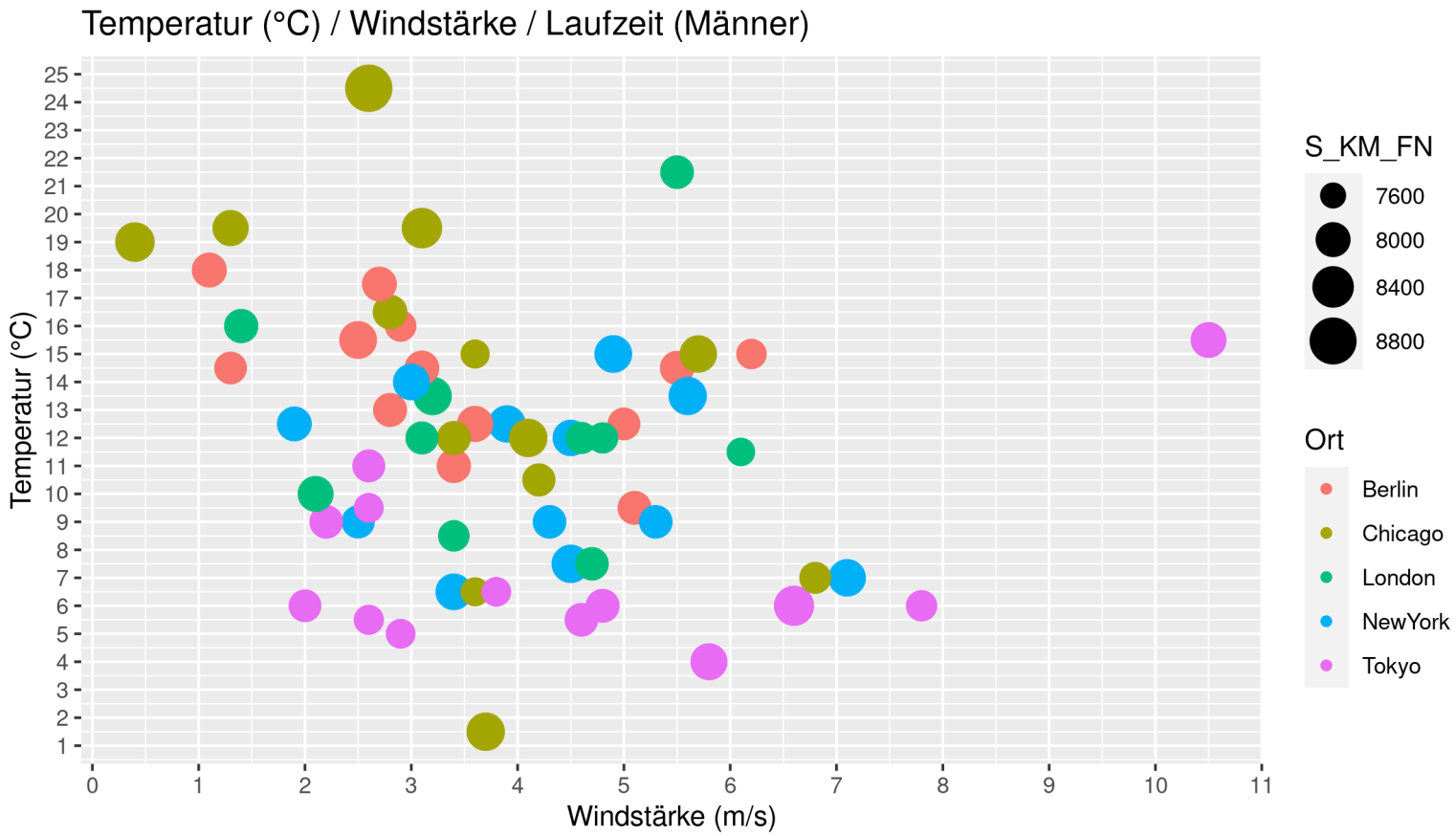
Marathondaten – Windstärke / Laufzeiten (Männer) (1)



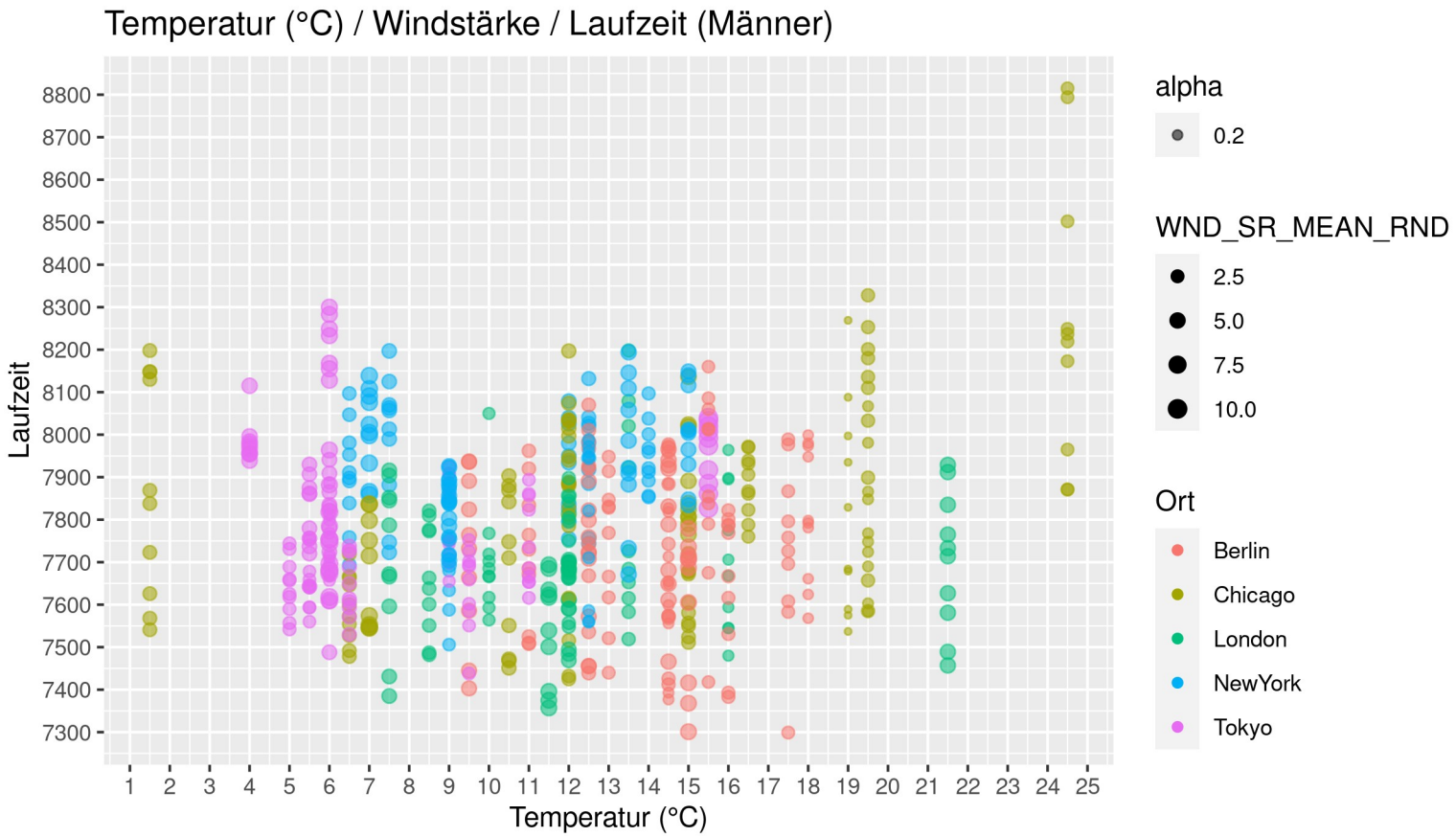
Marathondaten – Windstärke / Laufzeiten (Männer) (2)



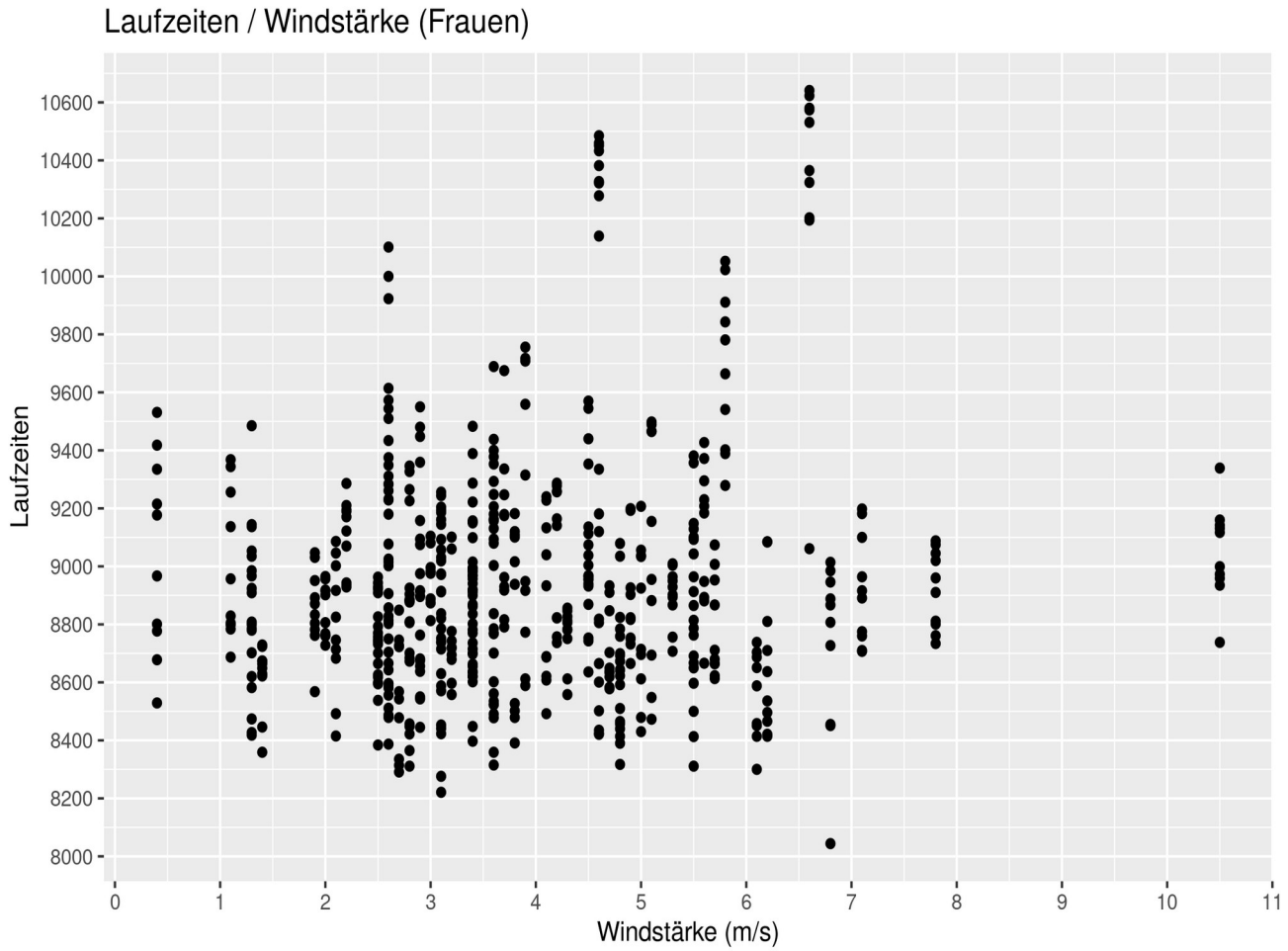
Marathondaten – Windstärke / Temperatur / Laufzeiten (Männer) (3)



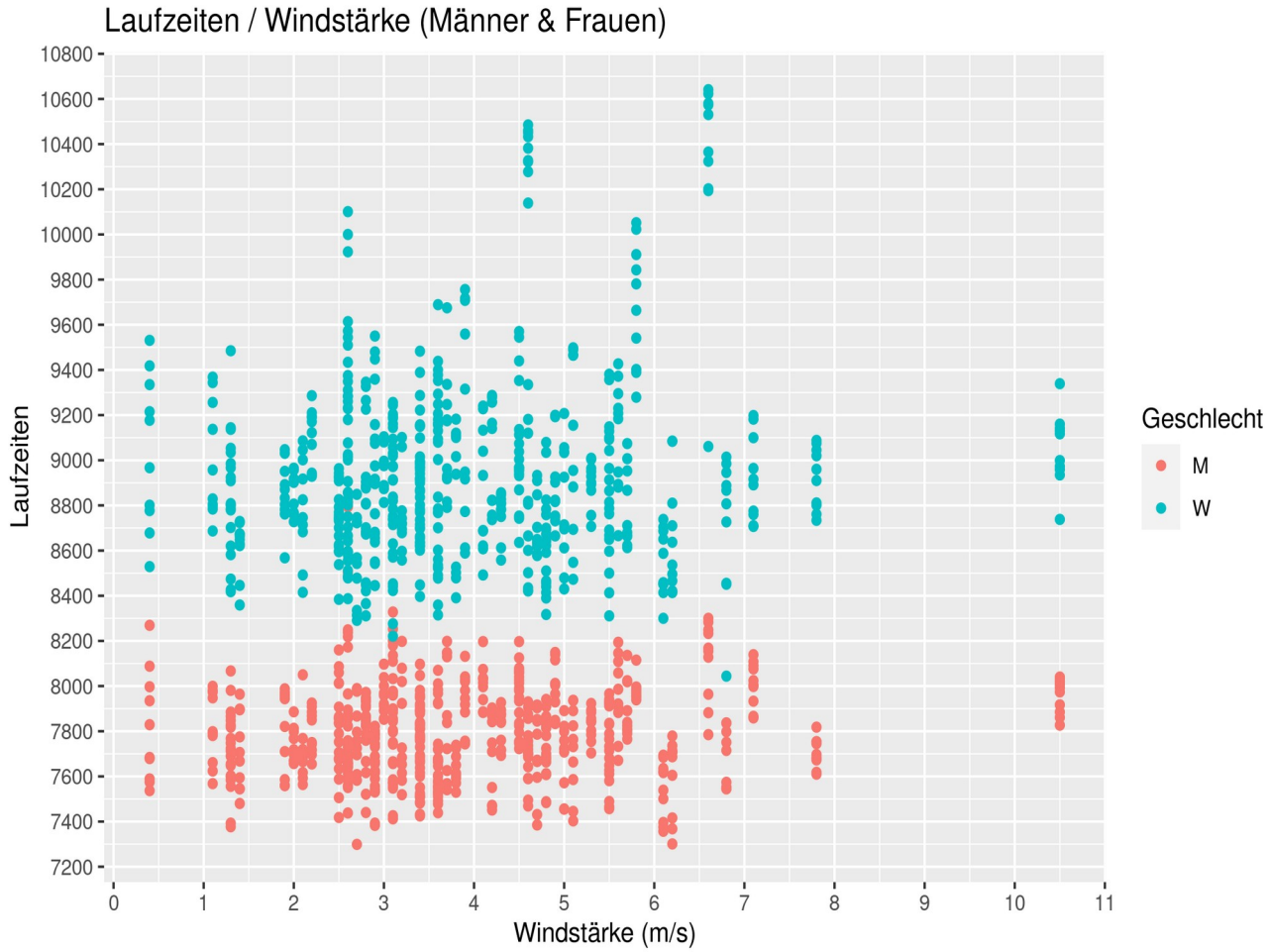
Marathondaten – Windstärke / Temperatur / Laufzeiten (Männer) (4)



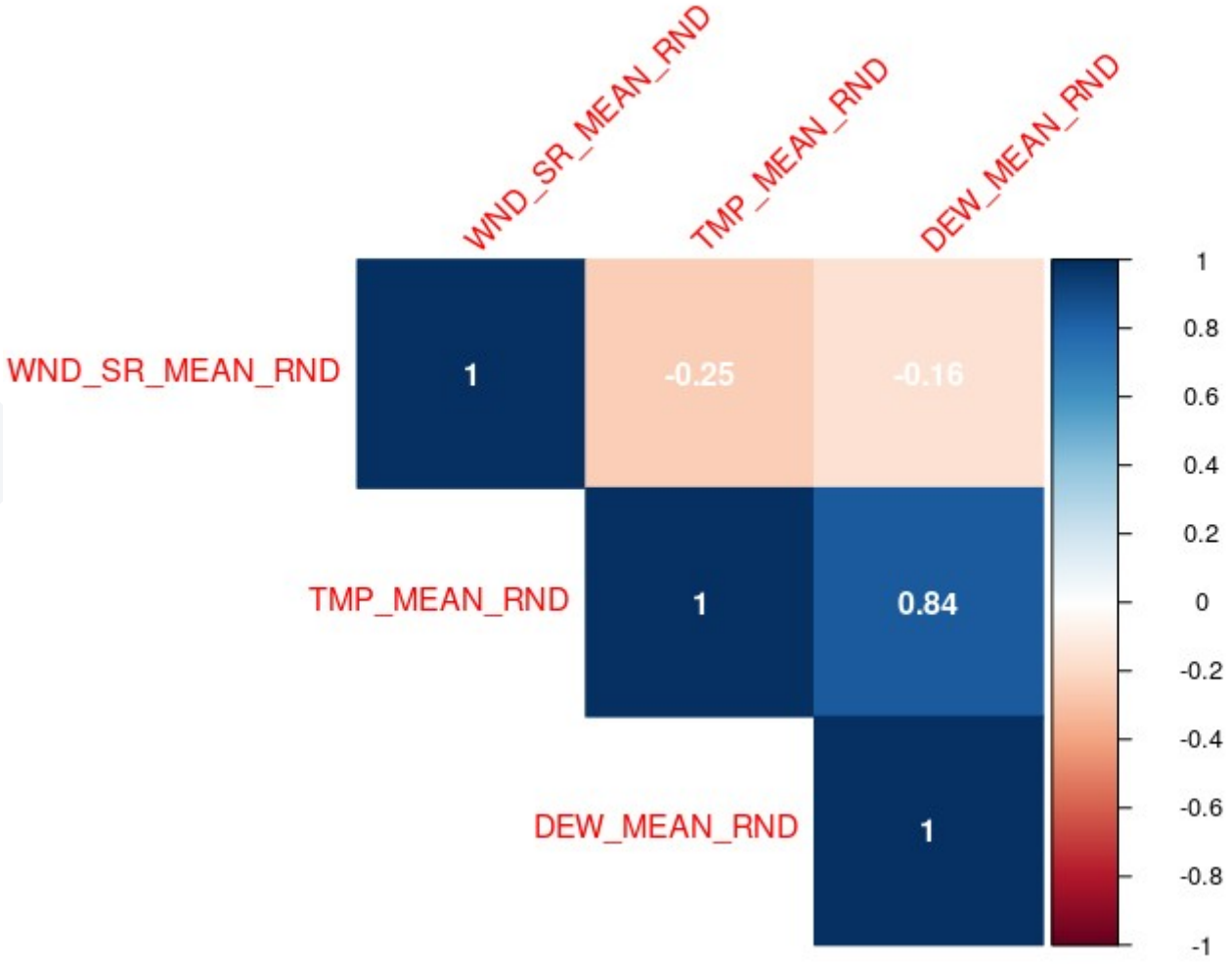
Marathondaten – Windstärke / Laufzeiten (Frauen)



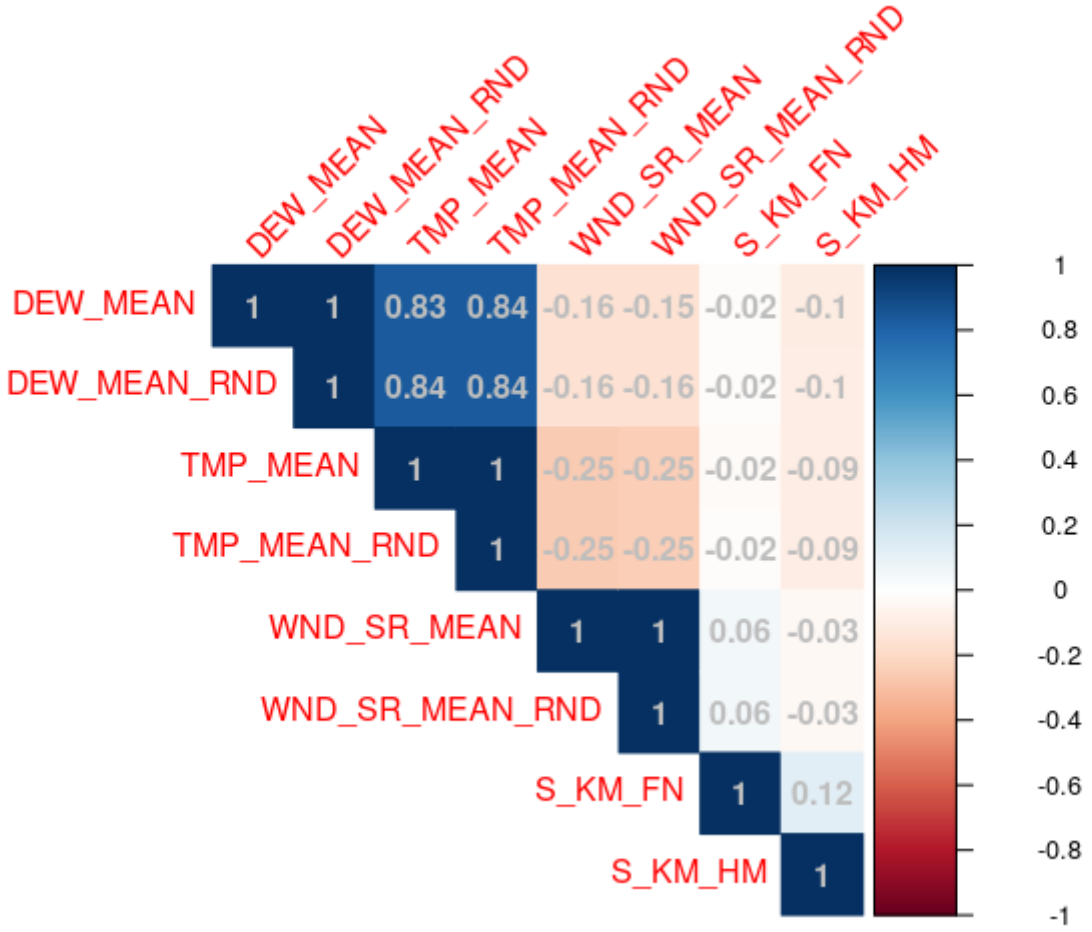
Marathondaten – Windstärke / Laufzeiten (Frauen & Männer)



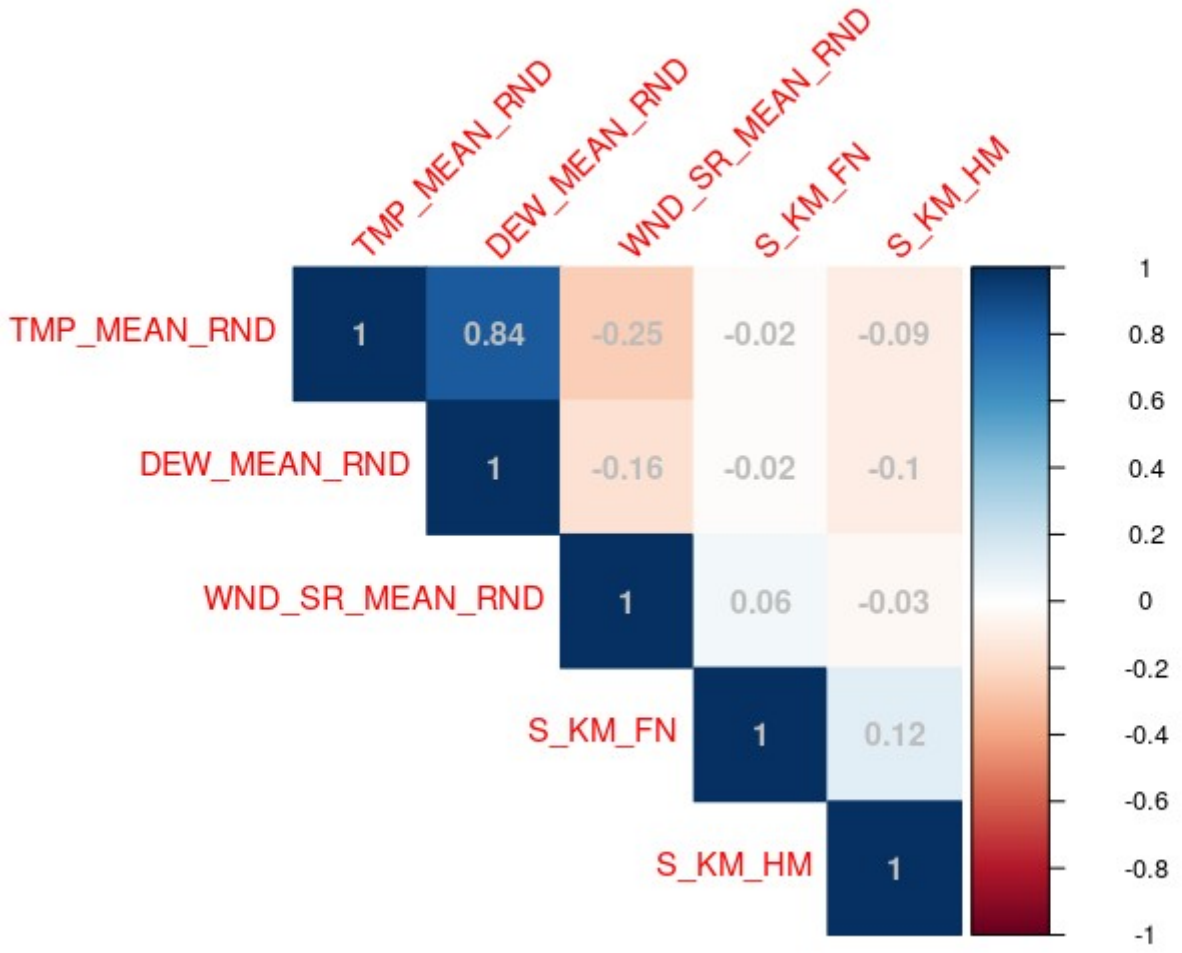
Korrelation (1)



Korrelation (2)



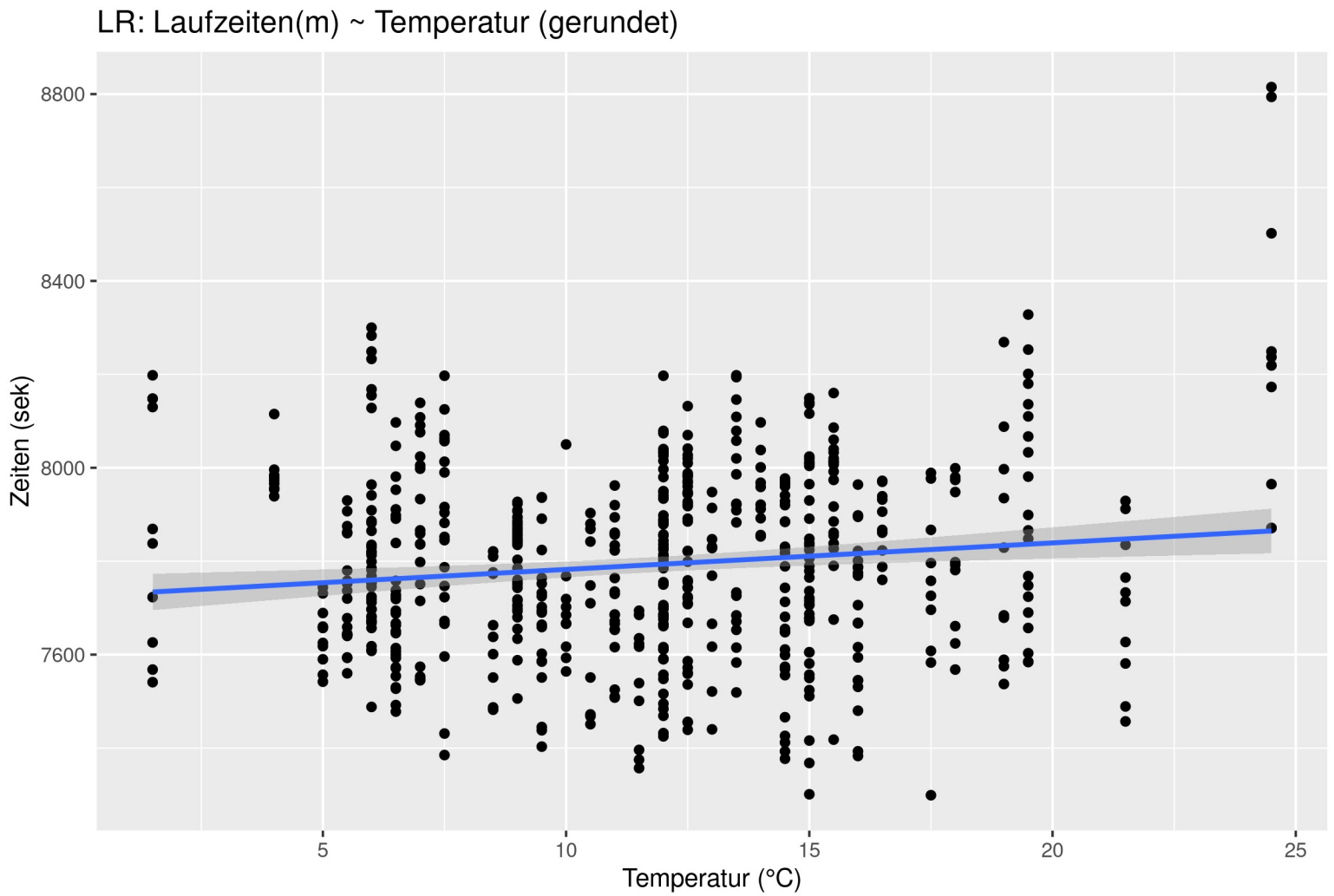
Korrelation (3)



Regression: Laufzeit (M) ~ Temperatur (gerundet) (1)

```
Call:
lm(formula = S_KM_FN ~ TMP_MEAN_RND, data = subset(df_wma_wetter_2,
    (Geschlecht == "M")))

Residuals:
    Min      1Q  Median      3Q     Max
-525.95 -133.97  -10.54  134.44  950.30

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7725.590     22.283 346.701  < 2e-16 ***
TMP_MEAN_RND    5.678      1.779   3.192  0.00148 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 203.6 on 608 degrees of freedom
Multiple R-squared:  0.01649,   Adjusted R-squared:  0.01487
F-statistic: 10.19 on 1 and 608 DF,  p-value: 0.001484
```

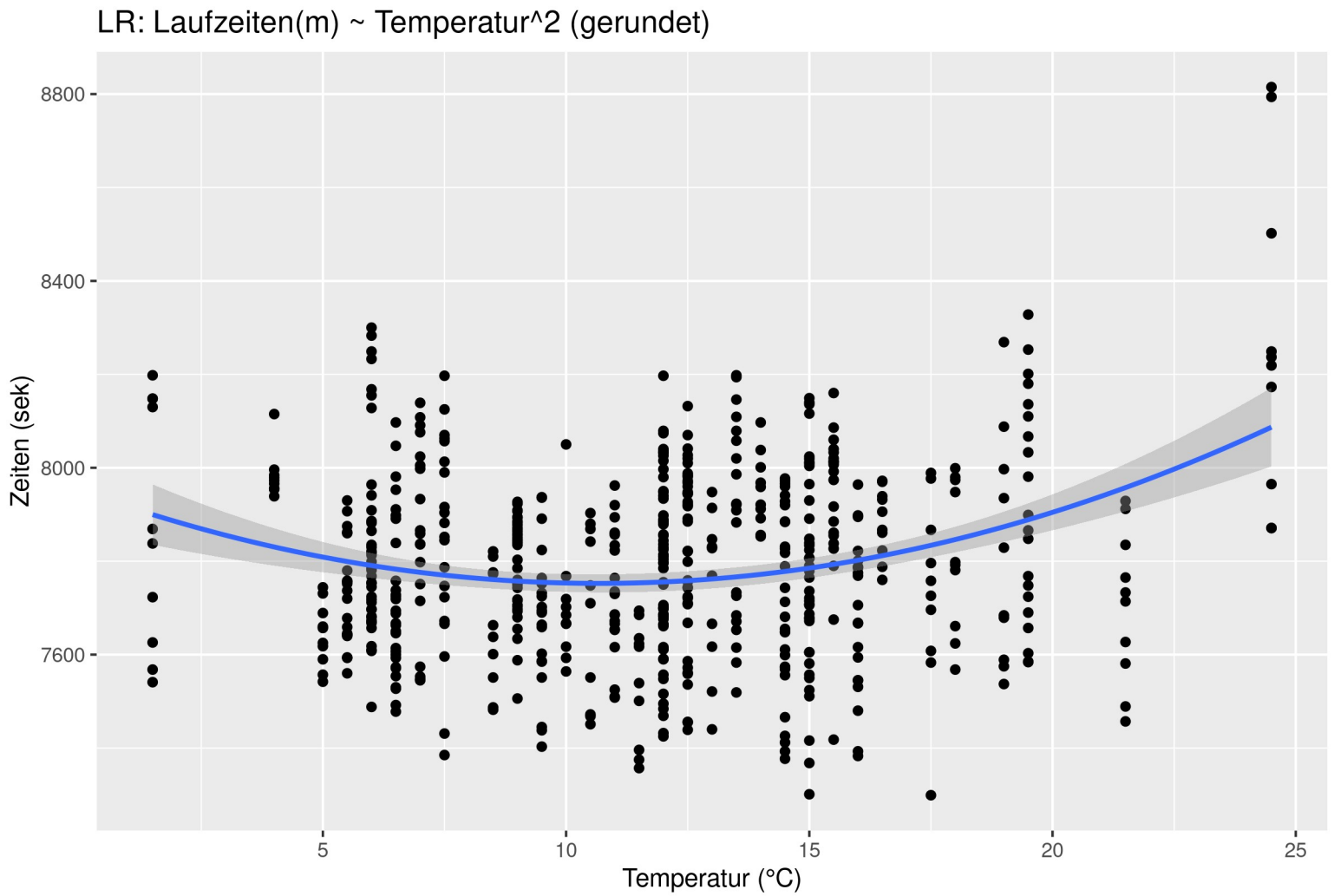
Regression: Laufzeit (M) ~ Temperatur (gerundet) (2)



Regression: Laufzeit (M) ~ Temperatur (gerundet) [poly-2] (1)

```
Call:
lm(formula = S_KM_FN ~ poly(TMP_MEAN_RND, 2), data = subset(df_wma_wetter_2,
    (Geschlecht == "M")))

Residuals:
    Min      1Q  Median      3Q     Max
-534.97 -139.14  -11.89  141.07  728.05

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            7791.677      8.001 973.781  < 2e-16 ***
poly(TMP_MEAN_RND, 2)1  650.009    197.622   3.289  0.00106 **
poly(TMP_MEAN_RND, 2)2 1225.325    197.622   6.200 1.04e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 197.6 on 607 degrees of freedom
Multiple R-squared:  0.07507,   Adjusted R-squared:  0.07202
F-statistic: 24.63 on 2 and 607 DF,  p-value: 5.184e-11
```

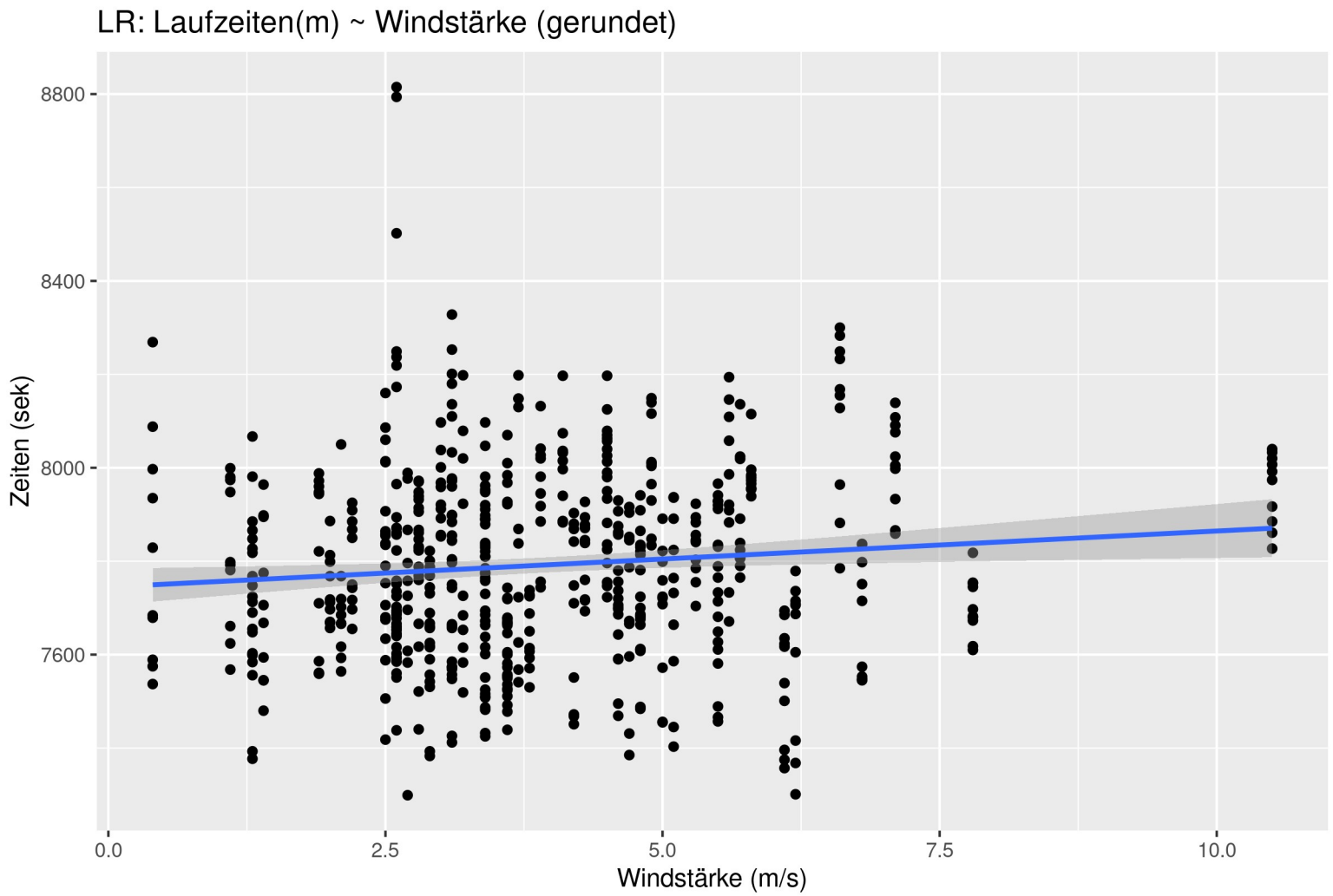
Regression: Laufzeit (M) ~ Temperatur (gerundet) [poly-2] (2)



Regression: Laufzeit (M) ~ Windstärke (gerundet) (1)

```
Call:
lm(formula = S_KM_FN ~ WND_SR_MEAN_RND, data = subset(df_wma_wetter_2,
    (Geschlecht == "M")))

Residuals:
    Min      1Q  Median      3Q     Max
-518.02 -133.86  -12.35  131.31 1039.15

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     7744.669     19.955 388.113  < 2e-16 ***
WND_SR_MEAN_RND   11.993      4.633   2.588  0.00987 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 204.2 on 608 degrees of freedom
Multiple R-squared:  0.0109,    Adjusted R-squared:  0.009272
F-statistic:   6.7 on 1 and 608 DF,  p-value: 0.009874
```

Regression: Laufzeit (M) ~ Windstärke (gerundet) (2)



Regression: Laufzeit (M) ~ Windstärke (gerundet) + Temperatur (gerundet) (1)

```
Call:
lm(formula = S_KM_FN ~ TMP_MEAN_RND + WND_SR_MEAN_RND, data = subset(df_wma_wetter_2,
    (Geschlecht == "M")))

Residuals:
    Min      1Q  Median      3Q     Max
-553.08 -132.84   -7.87  130.51  951.89

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    7641.830     32.427 235.660  < 2e-16 ***
TMP_MEAN_RND      7.264      1.819   3.994 7.29e-05 ***
WND_SR_MEAN_RND  16.659      4.724   3.526 0.000453 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 201.7 on 607 degrees of freedom
Multiple R-squared:  0.03623,   Adjusted R-squared:  0.03305
F-statistic: 11.41 on 2 and 607 DF,  p-value: 1.368e-05
```

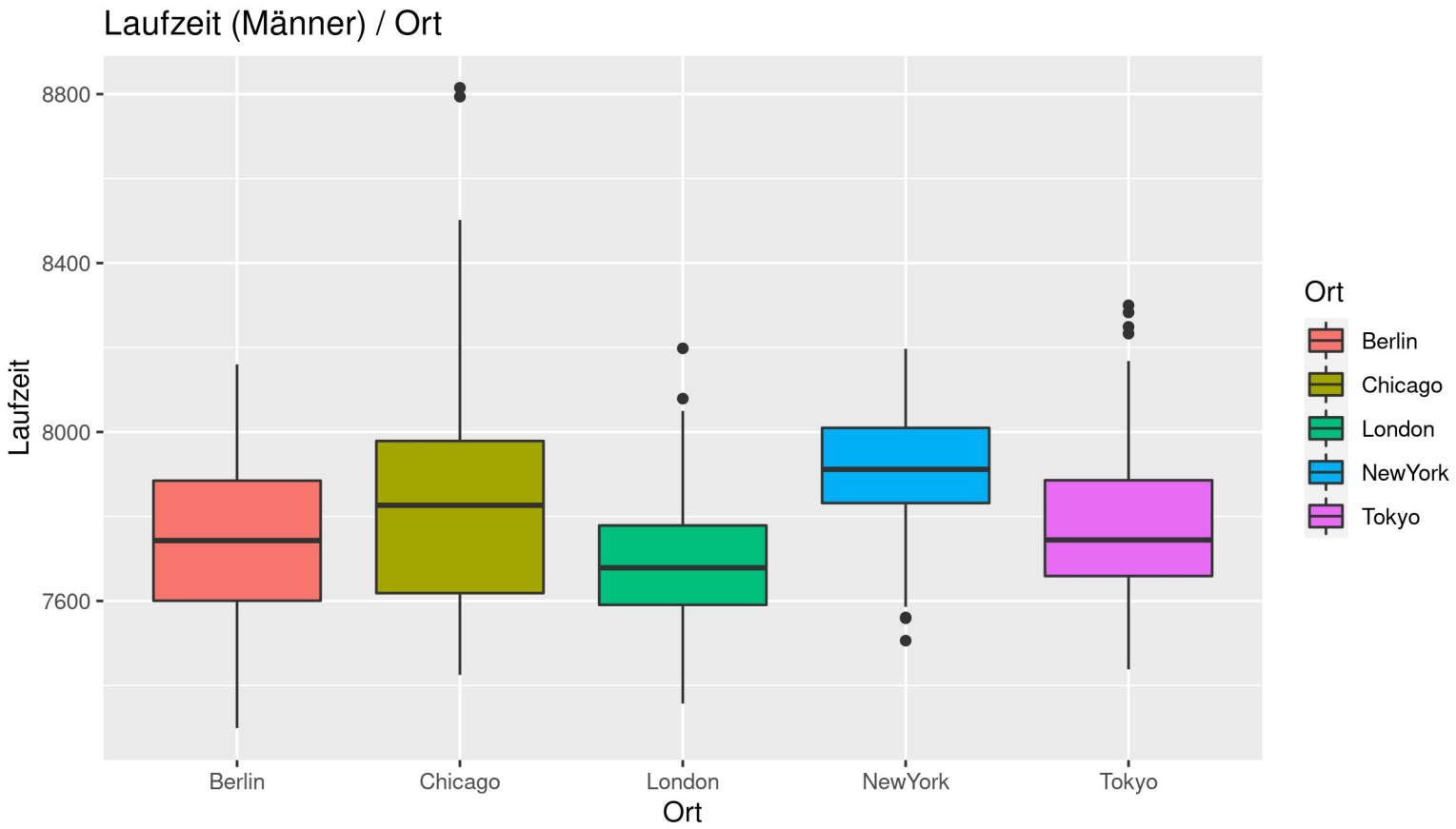
Regression: Laufzeit (M) ~ Windstärke (gerundet) + Temperatur [poly-2] (gerundet) (2)

```
Call:
lm(formula = S_KM_FN ~ poly(TMP_MEAN_RND, 2) + WND_SR_MEAN_RND,
    data = subset(df_wma_wetter_2, (Geschlecht == "M")))

Residuals:
    Min      1Q  Median      3Q     Max
-549.30 -129.62   -9.39  133.72  723.15

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            7720.892     19.612 393.692  < 2e-16 ***
poly(TMP_MEAN_RND, 2)1  846.825    201.568   4.201 3.05e-05 ***
poly(TMP_MEAN_RND, 2)2 1261.845    195.513   6.454 2.23e-10 ***
WND_SR_MEAN_RND          18.059      4.579   3.944 8.94e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 195.3 on 606 degrees of freedom
Multiple R-squared:  0.09822,   Adjusted R-squared:  0.09375
F-statistic:    22 on 3 and 606 DF,  p-value: 1.557e-13
```

Test – Statistiken (M)

```
> describeBy(df_wma_ort_skm_m$S_KM_FN, df_wma_ort_skm_m$Ort)

 Descriptive statistics by group
group: Berlin
   vars   n    mean     sd median trimmed    mad  min  max range  skew kurtosis   se
X1    1 130 7731.53 192.72   7743 7738.01 212.01 7299 8160   861 -0.23    -0.72 16.9
------------------------------------------------------------------------------------
group: Chicago
   vars   n    mean     sd median trimmed   mad  min  max range skew kurtosis    se
X1    1 130 7834.32 258.54 7826.5  7813.9 260.2 7425 8815  1390 0.95     1.53 22.68
------------------------------------------------------------------------------------
group: London
   vars   n    mean     sd median trimmed    mad  min  max range skew kurtosis    se
X1    1 100 7687.03 159.61 7678.5 7681.15 142.33 7357 8198   841 0.45      0.3 15.96
------------------------------------------------------------------------------------
group: NewYork
   vars   n    mean     sd median trimmed    mad  min  max range  skew kurtosis    se
X1    1 120 7905.24 147.04 7911.5 7910.36 137.88 7506 8197   691 -0.35    -0.28 13.42
------------------------------------------------------------------------------------
group: Tokyo
   vars   n    mean     sd median trimmed    mad  min  max range skew kurtosis    se
X1    1 130 7784.85 170.84 7744.5  7770.3 169.02 7438 8300   862  0.8     0.41 14.98
> |
```

Test – Statistiken (M) - Boxplot



Test – Levene (M)

```
> leveneTest(S_KM_FN ~ Ort, df_wma_ort_skm_m)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value    Pr(>F)
group   4  9.9712 8.078e-08 ***
      605
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Warnmeldung:
In leveneTest.default(y = y, group = group, ...) : group coerced to factor.
```

Test – One-Way-Test (M)

```
> oneway.test(S_KM_FN ~ Ort,  data = df_wma_ort_skm_m, var.equal = FALSE)

        One-way analysis of means (not assuming equal variances)

data:  S_KM_FN and Ort
F = 32.477, num df = 4.00, denom df = 298.77, p-value < 2.2e-16
```

Tests – Pairwise.t.test (M) (two.sided)

```
> pairwise.t.test(df_wma_ort_skm_m$S_KM_FN, df_wma_ort_skm_m$Ort,
+                 p.adjust.method = "bonferroni", alternative = "two.sided",
+                 paired = FALSE, pool.sd = FALSE)

        Pairwise comparisons using t tests with non-pooled SD

data:  df_wma_ort_skm_m$S_KM_FN and df_wma_ort_skm_m$Ort

        Berlin  Chicago London  NewYork
Chicago 0.00341 -       -       -
London  0.56857 2.7e-06 -       -
NewYork 3.9e-13 0.07697 < 2e-16 -
Tokyo   0.19014 0.70034 0.00013 7.6e-08

P value adjustment method: bonferroni
```

Tests – Pairwise.t.test (M) (greater / less)

```
> pairwise.t.test(df_wma_ort_skm_m$S_KM_FN, df_wma_ort_skm_m$Ort,
+                 p.adjust.method = "bonferroni", alternative = "greater",
+                 paired = FALSE, pool.sd = FALSE)

        Pairwise comparisons using t tests with non-pooled SD

data:  df_wma_ort_skm_m$S_KM_FN and df_wma_ort_skm_m$Ort

        Berlin  Chicago London  NewYork
Chicago 0.0017  -       -       -
London  1.0000  1.0000  -       -
NewYork 2.0e-13 0.0385  < 2e-16 -
Tokyo   0.0951  1.0000  6.3e-05 1.0000

P value adjustment method: bonferroni
> # pairwise-test
> pairwise.t.test(df_wma_ort_skm_m$S_KM_FN, df_wma_ort_skm_m$Ort,
+                 p.adjust.method = "bonferroni", alternative = "less",
+                 paired = FALSE, pool.sd = FALSE)

        Pairwise comparisons using t tests with non-pooled SD

data:  df_wma_ort_skm_m$S_KM_FN and df_wma_ort_skm_m$Ort

        Berlin Chicago London NewYork
Chicago 1.00   -       -      -
London  0.28   1.3e-06 -      -
NewYork 1.00   1.00    1.00   -
Tokyo   1.00   0.35    1.00   3.8e-08

P value adjustment method: bonferroni
```

Tests – t-Test (M) (Berlin - NewYork)

```
> t.test(x=subset(df_wma_ort_skm_m, (Ort=='Berlin'), select = c(S_KM_FN)),
+        y=subset(df_wma_ort_skm_m, (Ort=='NewYork'), select = c(S_KM_FN)),
+        paired = FALSE, conf.level = 0.95, var.equal = FALSE, alternative = "greater"
+ )

Welch Two Sample t-test

data:
subset(df_wma_ort_skm_m, (Ort == "Berlin"), select = c(S_KM_FN)) and
subset(df_wma_ort_skm_m, (Ort == "NewYork"), select = c(S_KM_FN))

t = -8.048, df = 239.68, p-value = 1
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -209.3519       Inf
sample estimates:
mean of x mean of y
 7731.531  7905.242
```

Tests – t-Test (M) (Berlin - London)

```
> t.test(x=subset(df_wma_ort_skm_m, (Ort=='Berlin'), select = c(S_KM_FN)),
+        y=subset(df_wma_ort_skm_m, (Ort=='London'), select = c(S_KM_FN)),
+        paired = FALSE, conf.level = 0.95, var.equal = FALSE, alternative = "greater"
+ )

Welch Two Sample t-test

data:
subset(df_wma_ort_skm_m, (Ort == "Berlin"), select = c(S_KM_FN)) and
subset(df_wma_ort_skm_m, (Ort == "London"), select = c(S_KM_FN))

t = 1.9142, df = 226.73, p-value = 0.02843
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 6.104252      Inf
sample estimates:
mean of x mean of y
 7731.531  7687.030
```

Tests – t-Test (M) (NewYork - Chicago)

```
> t.test(x=subset(df_wma_ort_skm_m, (Ort=='NewYork'), select = c(S_KM_FN)),
+        y=subset(df_wma_ort_skm_m, (Ort=='Chicago'), select = c(S_KM_FN)),
+        paired = FALSE, conf.level = 0.95, var.equal = FALSE, alternative = "greater"
+ )

Welch Two Sample t-test

data:
subset(df_wma_ort_skm_m, (Ort == "NewYork"), select = c(S_KM_FN)) and
subset(df_wma_ort_skm_m, (Ort == "Chicago"), select = c(S_KM_FN))

t = 2.6913, df = 207.61, p-value = 0.003849
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 27.38095      Inf
sample estimates:
mean of x mean of y
 7905.242  7834.323
```