Lineare Regression

Jan-Philipp Kolb

20 Juni 2017

Die lineare Regression

Maindonald - Data Analysis

- EinfÃ¼hrung in R
- Datenanalyse
- Statistische Modelle
- Inferenzkonzepte
- Regression mit einem PrÃ¤diktor
- Multiple lineare Regression
- Ausweitung des linearen Modells
- ...

Lineare Regression in R - Beispieldatensatz

John H. Maindonald and W. John Braun

DAAG - Data Analysis and Graphics Data and Functions

```
install.packages("DAAG")
```

```
library("DAAG")
data(roller)
```

help on roller data:

```
?roller
```

Das lineare Regressionsmodell in R

SchÃ¤tzen eines Regressionsmodells:

```
roller.lm <- lm(depression ~ weight, data = roller)
```

So bekommt man die SchÃ¤tzwerte:

```
summary(roller.lm)
```

```
##
## Call:
## lm(formula = depression ~ weight, data = roller)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.180 -5.580 -1.346  5.920  8.020
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.0871     4.7543  -0.439  0.67227
```

Summary des Modells

```
summary(roller.lm)
```

```
##
## Call:
## lm(formula = depression ~ weight, data = roller)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.180 -5.580 -1.346  5.920  8.020
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.0871     4.7543  -0.439  0.67227
## weight        2.6667     0.7002   3.808  0.00518 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 6.735 on 8 degrees of freedom
```

R arbeitet mit Objekten

- `roller.lm` ist nun ein spezielles Regressions-Objekt
- Auf dieses Objekt kÃ¶nnen nun verschiedene Funktionen angewendet werden

```
predict(roller.lm) # Vorhersage
```

```
##         1         2         3         4         5
##  2.979669  6.179765  6.713114 10.713233 12.046606 14.180
##         8         9        10
## 18.180121 24.046962 30.980502
```

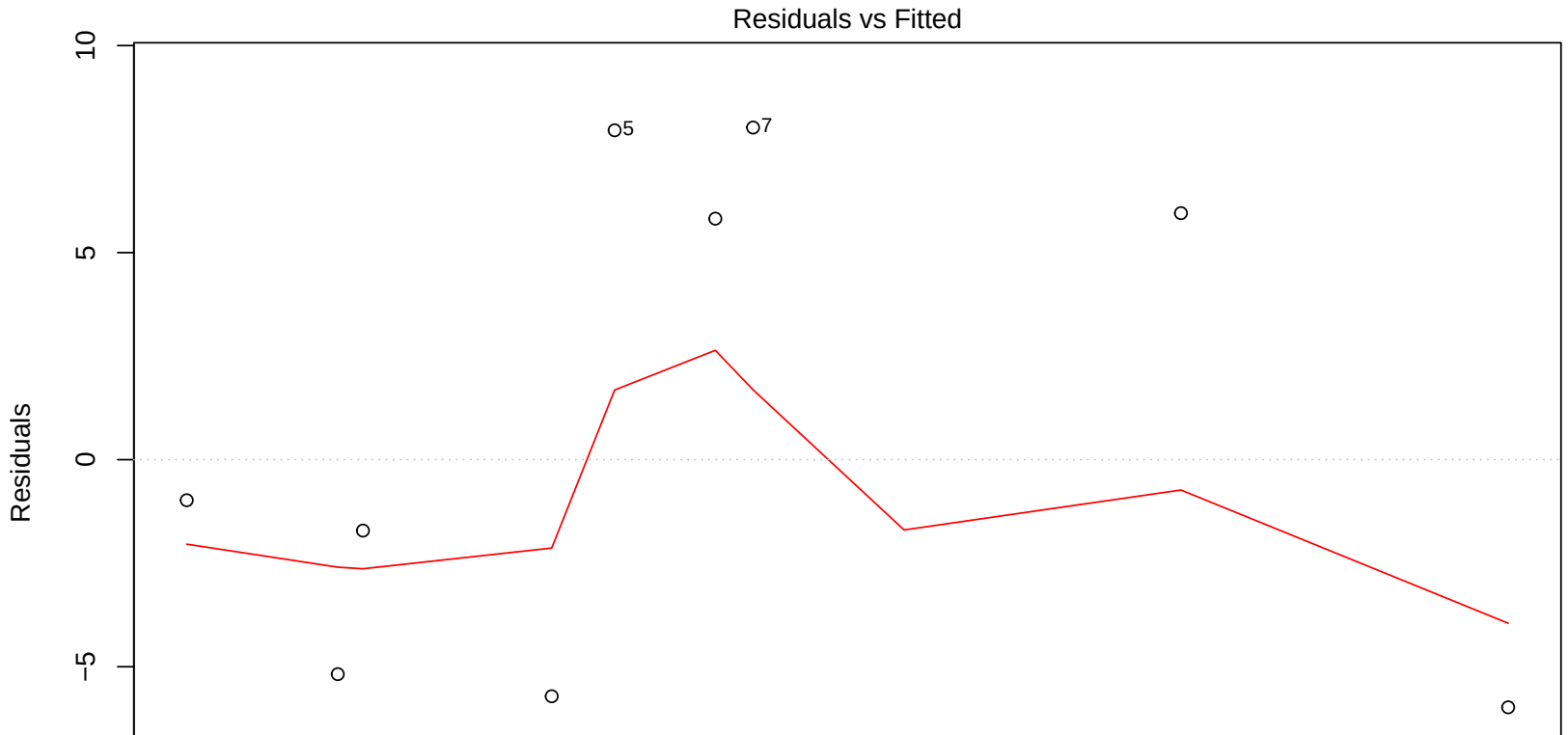
```
resid(roller.lm) # Residuen
```

```
##          1          2          3          4          5
## -0.9796695 -5.1797646 -1.7131138 -5.7132327  7.9533944
##          7          8          9         10
##  8.0199738 -8.1801213  5.9530377 -5.9805017
```

Residuenplot

- Sind Annahmen des linearen Regressionsmodells verletzt?
- Dies ist der Fall, wenn ein Muster abweichend von einer Linie zu erkennen ist.
- Hier ist der Datensatz sehr klein

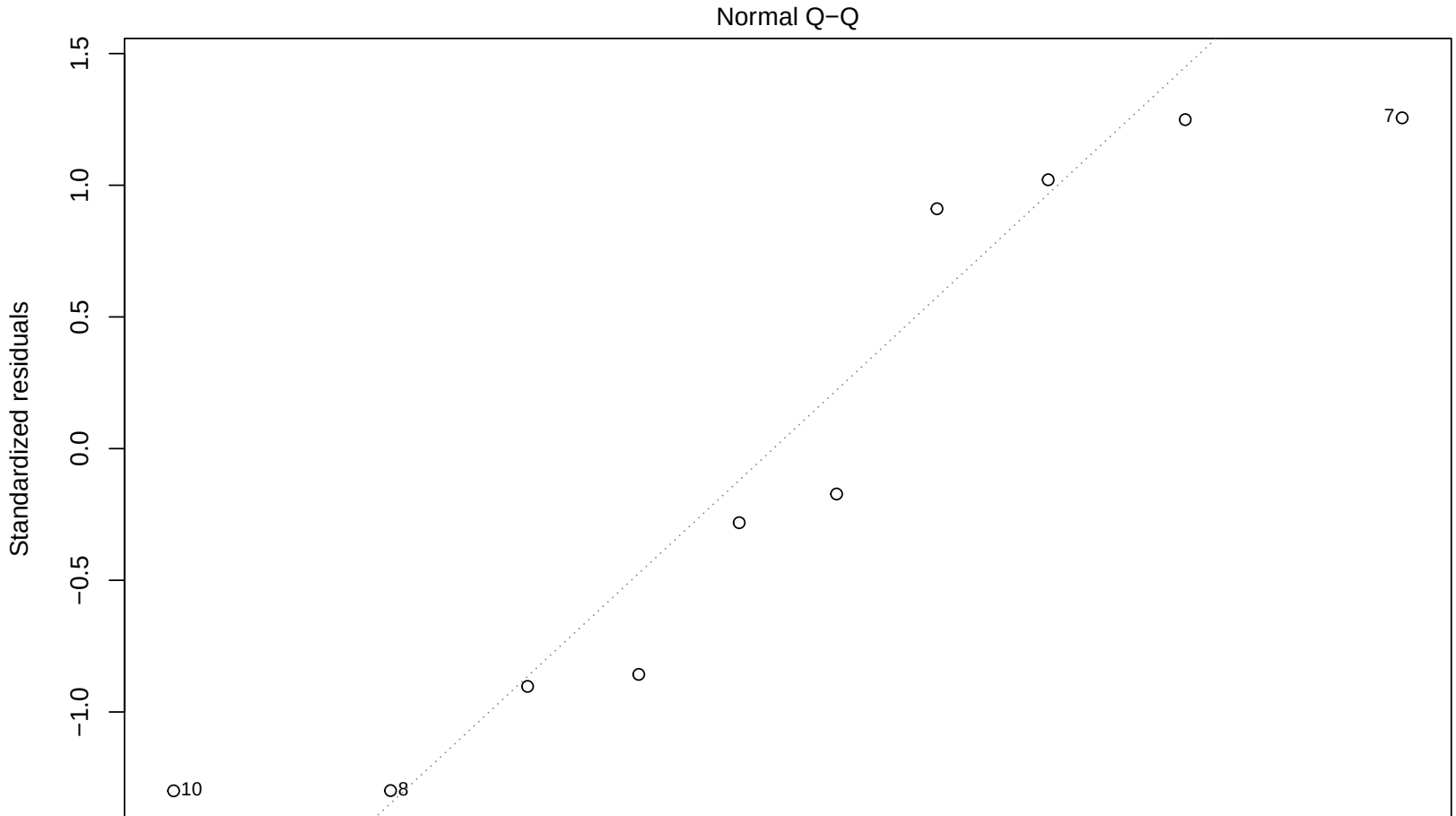
```
plot(roller.lm,1)
```



Residuenplot

- Wenn die Residuen normalverteilt sind sollten sie auf einer Linie liegen.

```
plot(roller.lm,2)
```



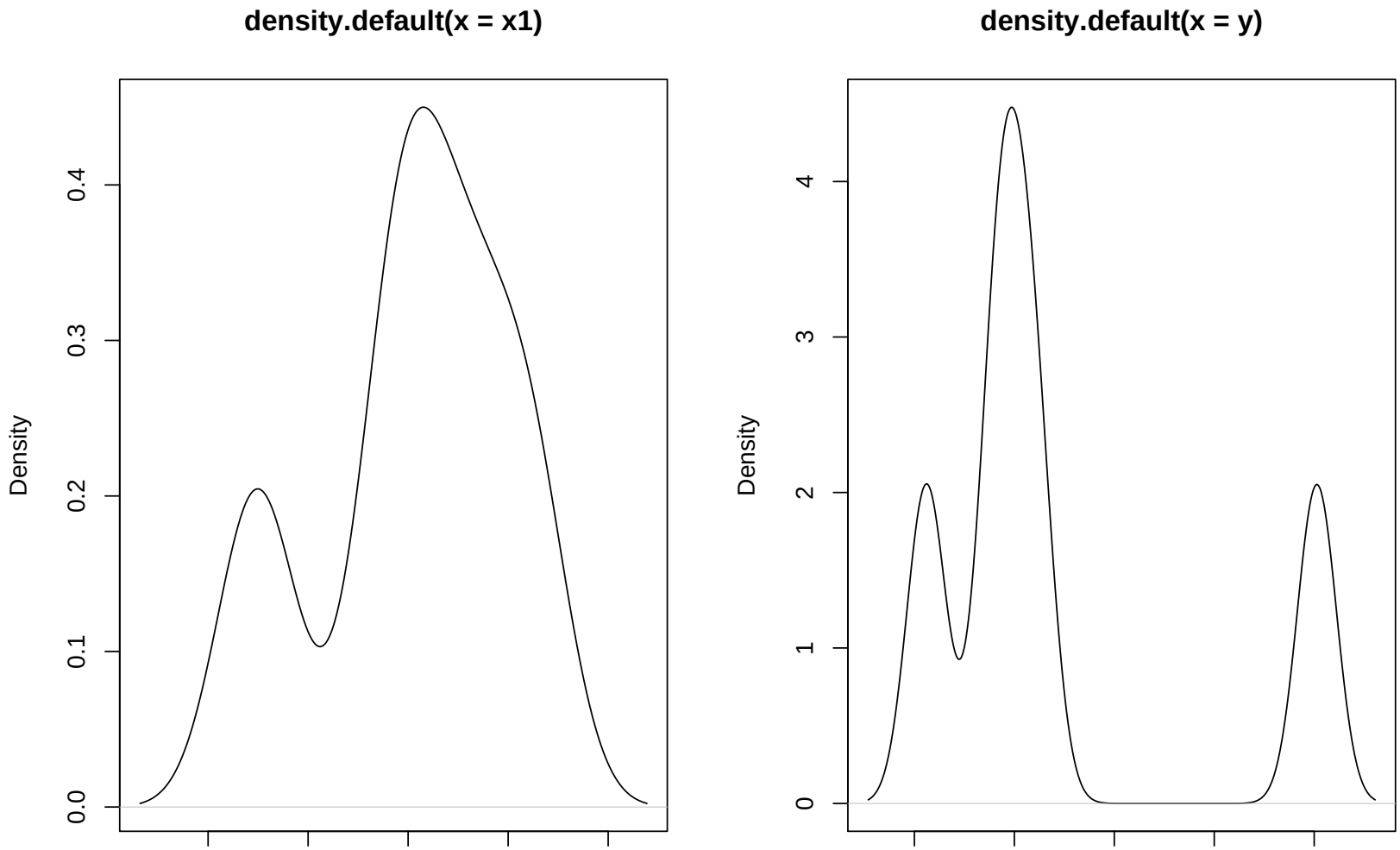
Regressionsdiagnostik mit Basis-R

Ein einfaches Modell

```
N <- 5
x1 <- rnorm(N)
y <- runif(N)
```

Die Dichte der beiden Vektoren

```
par(mfrow=c(1,2))
plot(density(x1))
plot(density(y))
```



Modellvorhersage machen

```
mod1 <- lm(y~x1)
pre <- predict(mod1)
y
```

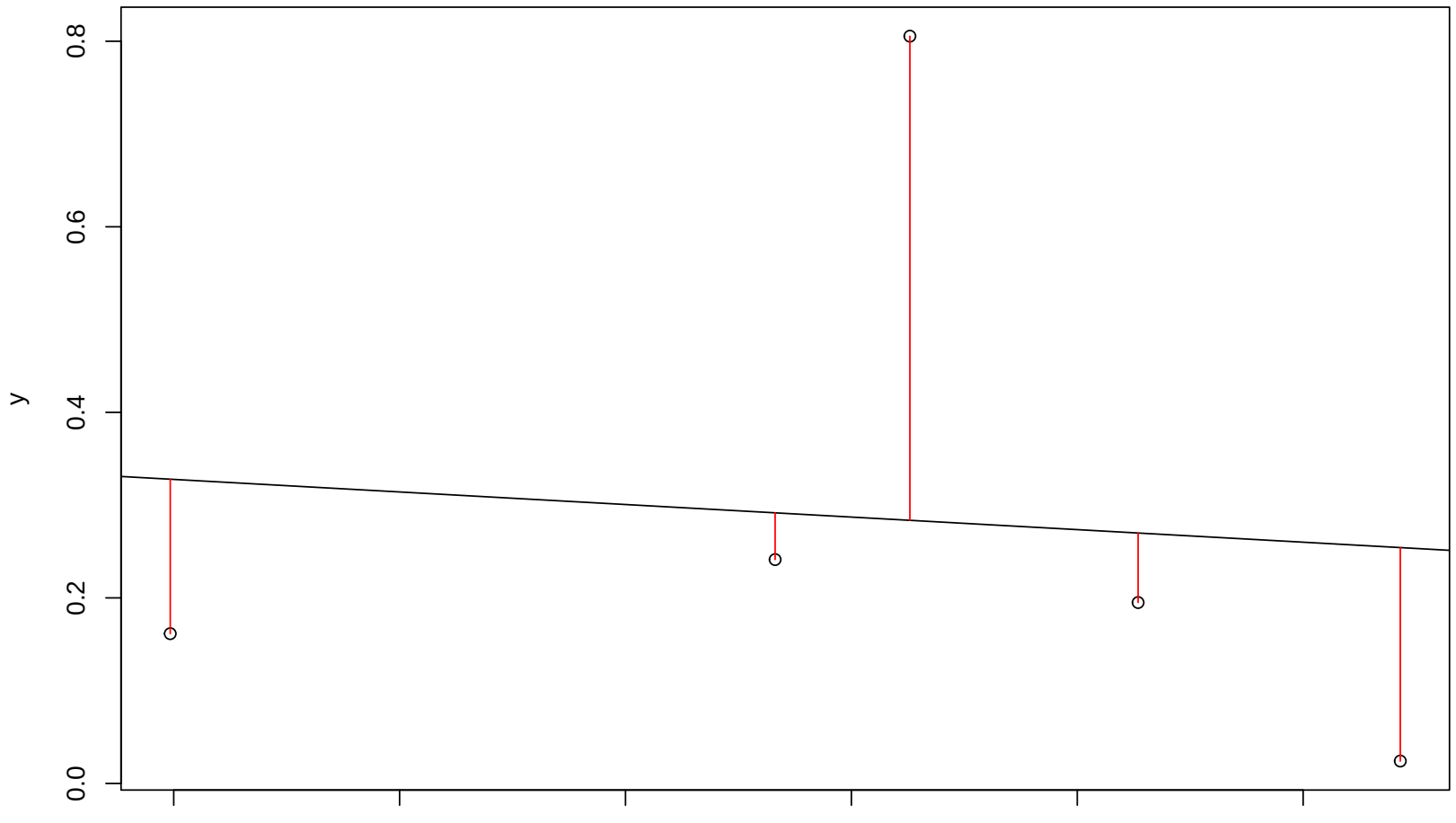
```
## [1] 0.80546933 0.19499632 0.24129626 0.02413855 0.161443
```

```
pre
```

```
##         1         2         3         4         5
## 0.2836019 0.2699229 0.2916785 0.2542092 0.3279318
```

Regressionsdiagnostik mit Basis-R

```
plot(x1,y)
abline(mod1)
segments(x1, y, x1, pre, col="red")
```



Beispieldaten LuftqualitÃ¤t

```
library(datasets)
?airquality
```

airquality {datasets}

New York Air Quality Measurements

Description

Daily air quality measurements in New York, May to September 1973.

Usage

```
airquality
```

Format

A data frame with 154 observations on 6 variables.

[,1] Ozone numeric Ozone (ppb)

[,2] Solar.R numeric Solar R (lang)

Das visreg-Paket

Ein Modell wird auf dem airquality Datensatz geschÃ¤tzt

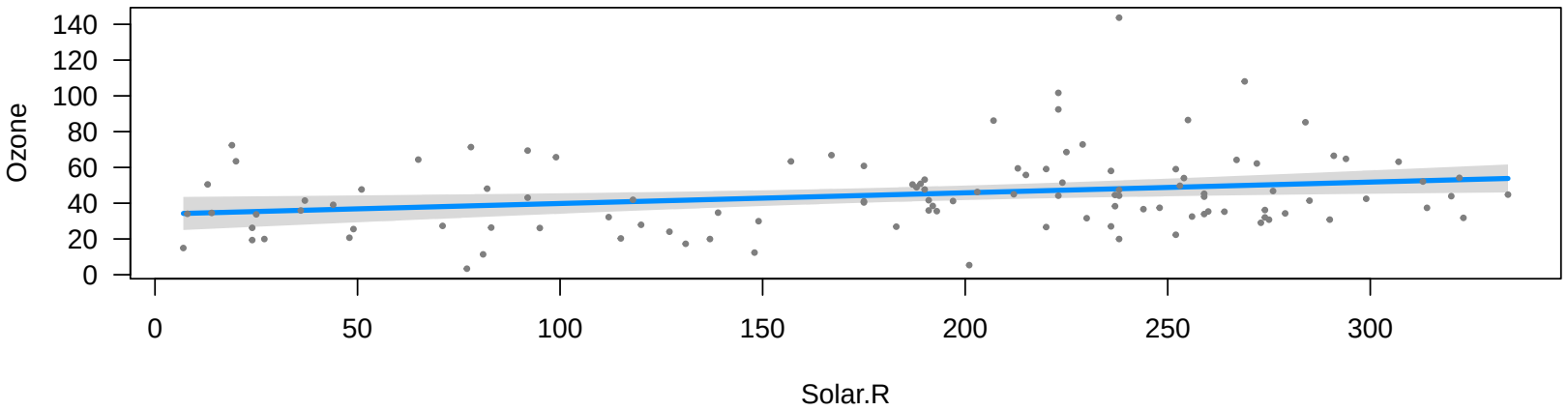
```
install.packages("visreg")
```

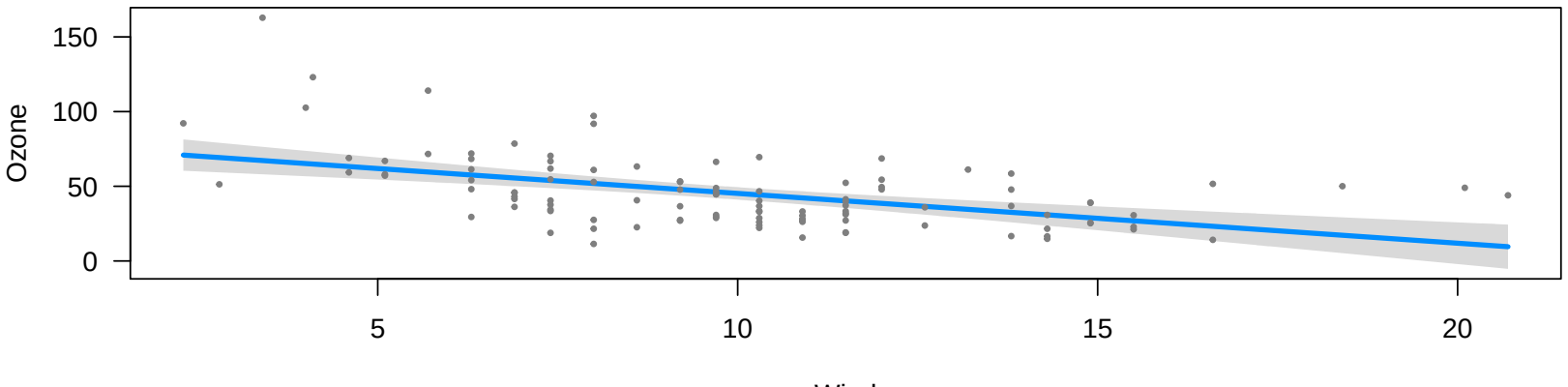
```
library(visreg)
fit <- lm(Ozone ~ Solar.R + Wind + Temp, data = airquality)
summary(fit)
```

```
##
## Call:
## lm(formula = Ozone ~ Solar.R + Wind + Temp, data = airqu
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.485 -14.219  -3.551  10.097  95.619
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

Visualisierung

```
par(mfrow=c(2,1))
visreg(fit)
```

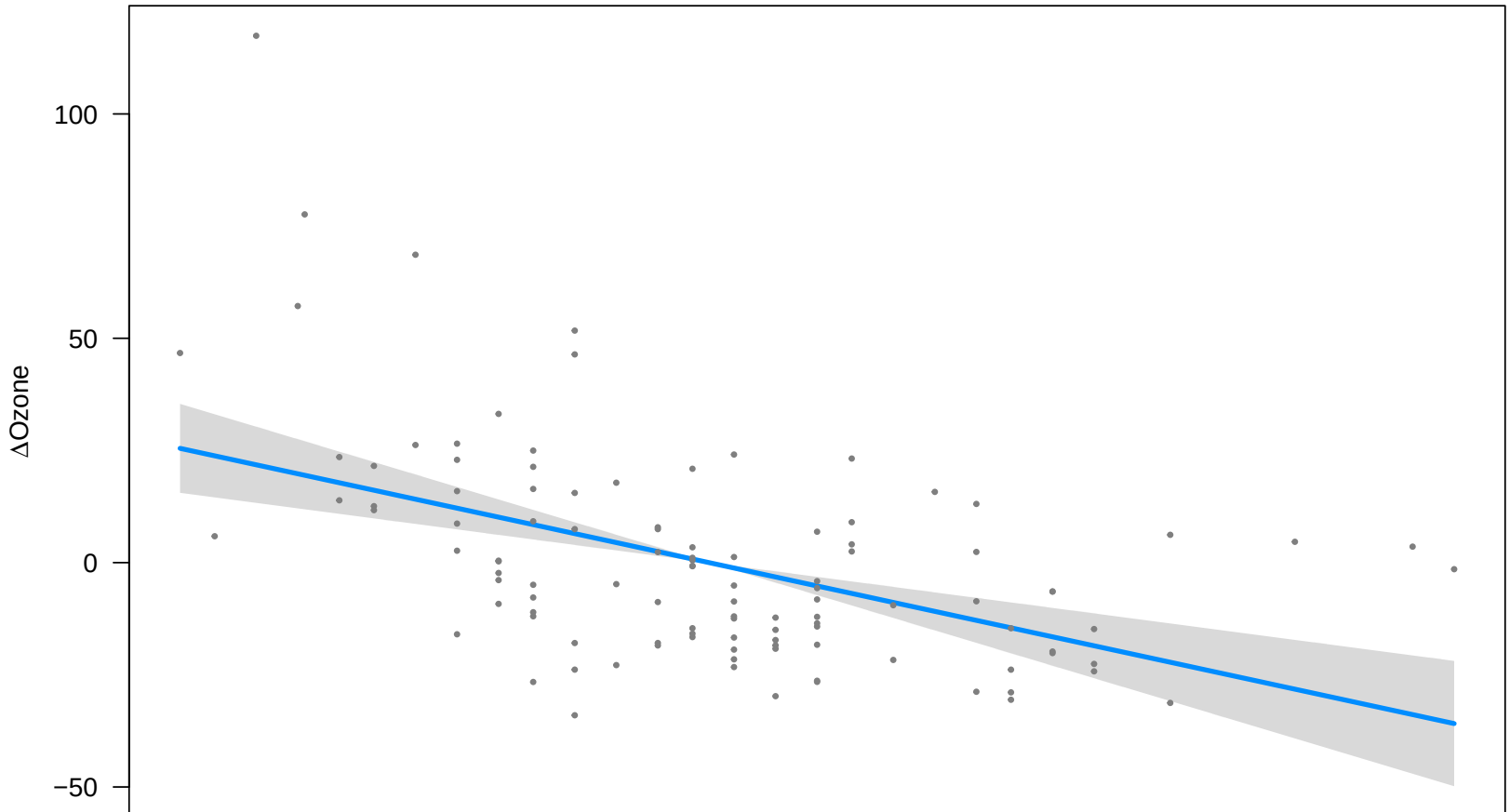




Und dann mit `visreg` visualisiert.

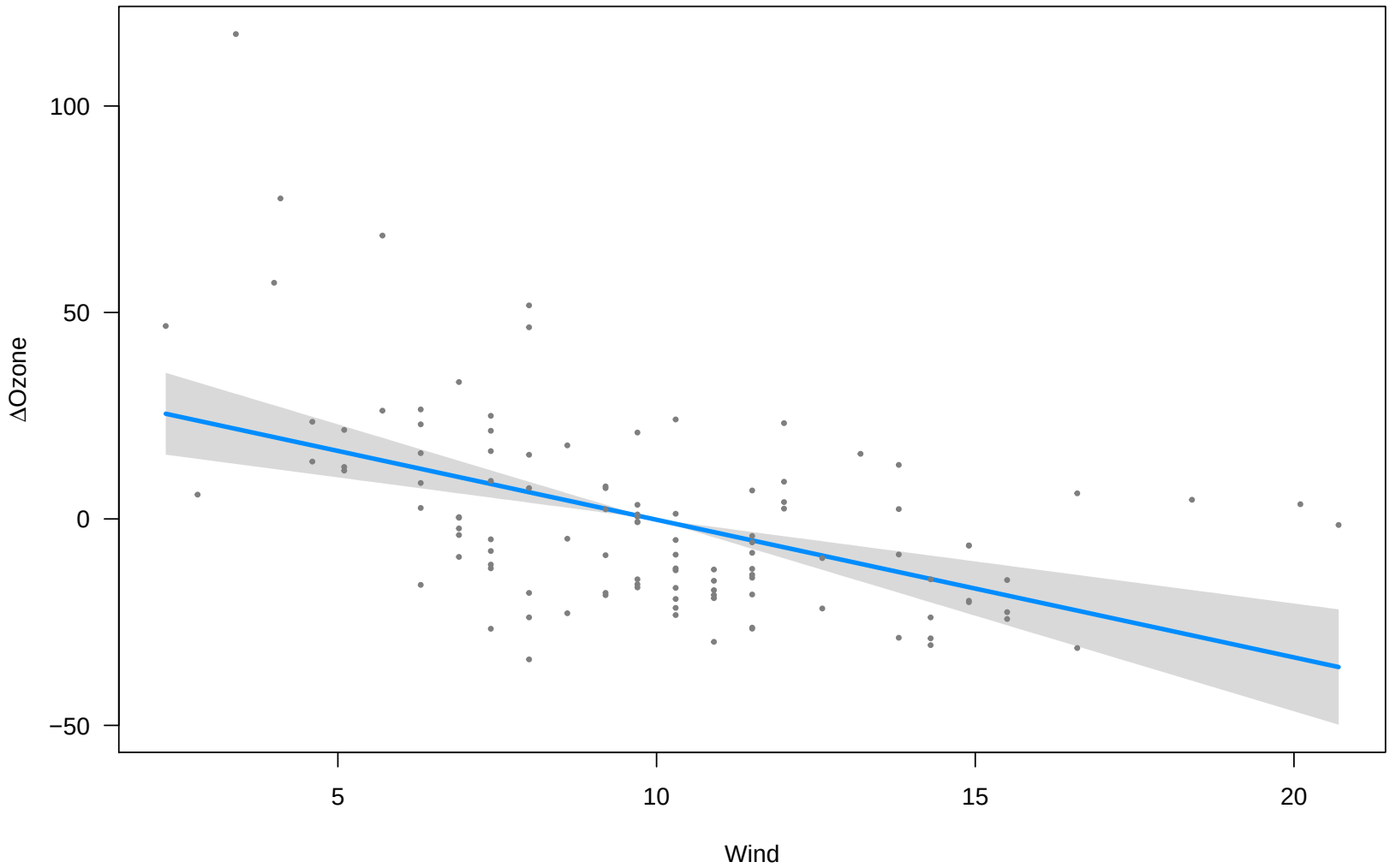
- Zweites Argument - Spezifikation erklÃ¤rende Variable fÃ¼r Visualisierung

```
visreg(fit, "Wind", type = "contrast")
```



Visualisierung mit dem Paket `visreg`

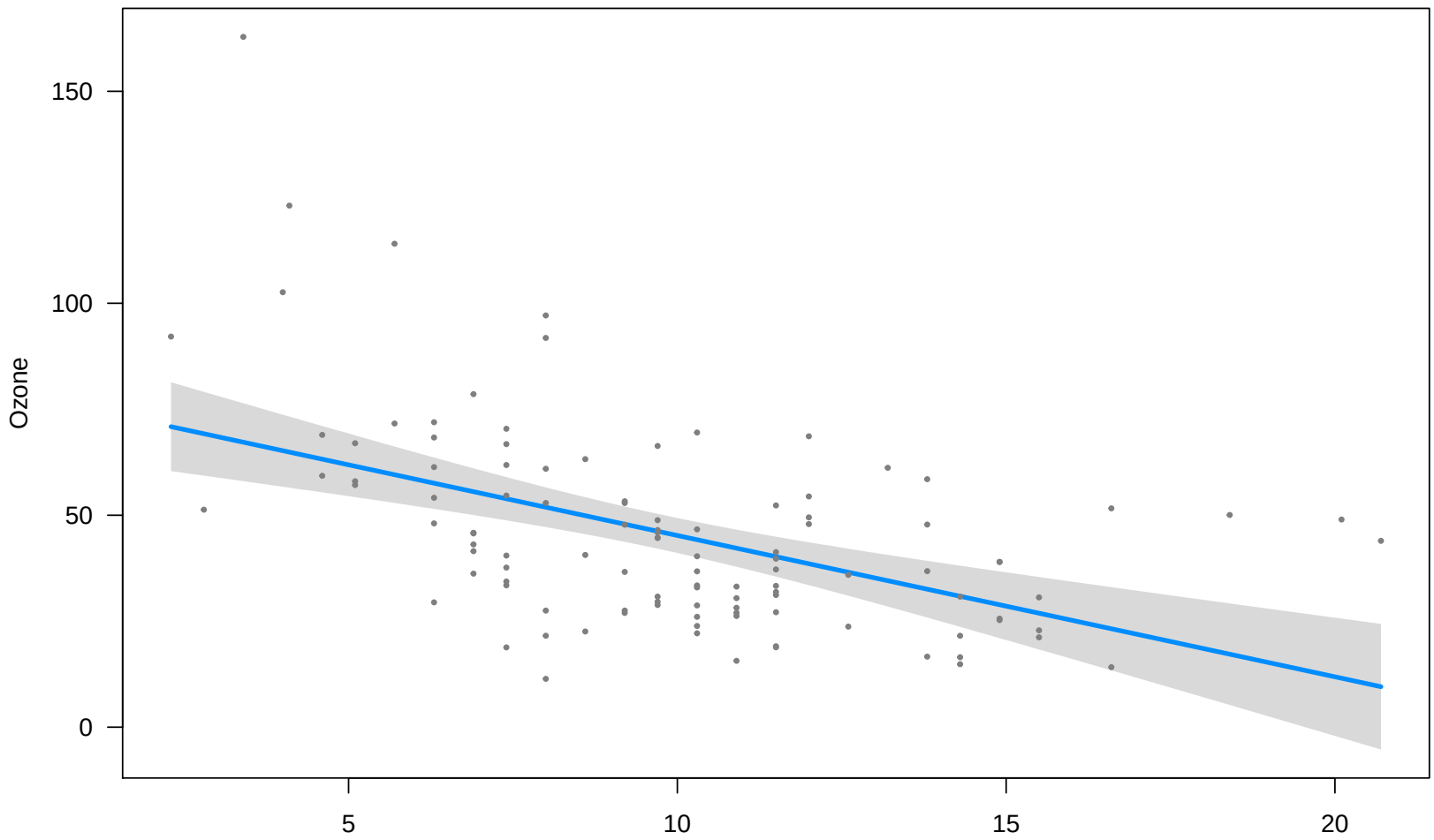
```
visreg(fit, "Wind", type = "contrast")
```



Das visreg-Paket

- Das Default-Argument fÃ¼r type ist conditional.

```
visreg(fit, "Wind", type = "conditional")
```



Regression mit Faktoren

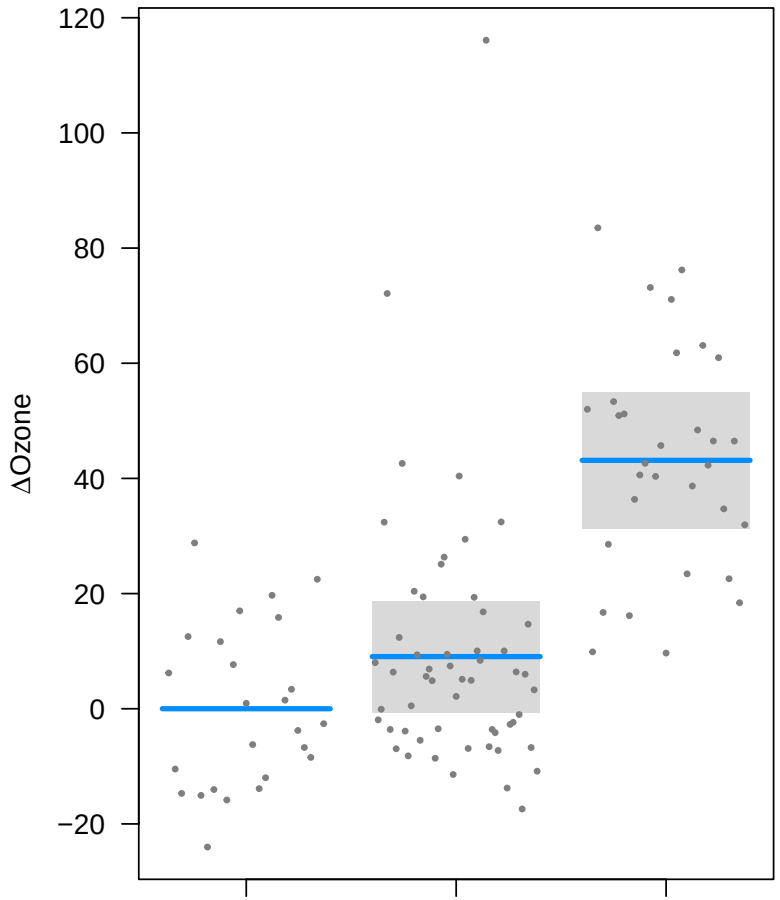
Mit visreg kÃ¶nnen die Effekte bei Faktoren visualisiert werden.

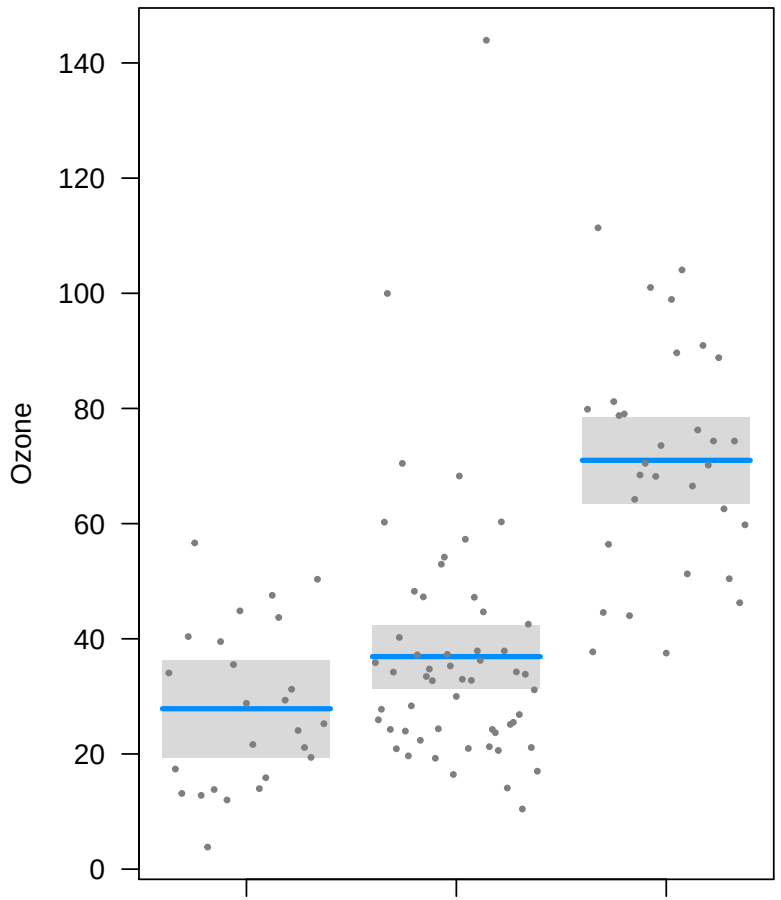
```
airquality$Heat <- cut(airquality$Temp, 3,
    labels=c("Cool", "Mild", "Hot"))
fit.heat <- lm(Ozone ~ Solar.R + Wind + Heat,
    data = airquality)
summary(fit.heat)
```

```
##
## Call:
## lm(formula = Ozone ~ Solar.R + Wind + Heat, data = airqu
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.473 -12.794  -2.686   8.461 107.035
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

Effekte von Faktoren

```
par(mfrow=c(1,2))
visreg(fit.heat, "Heat", type = "contrast")
visreg(fit.heat, "Heat", type = "conditional")
```





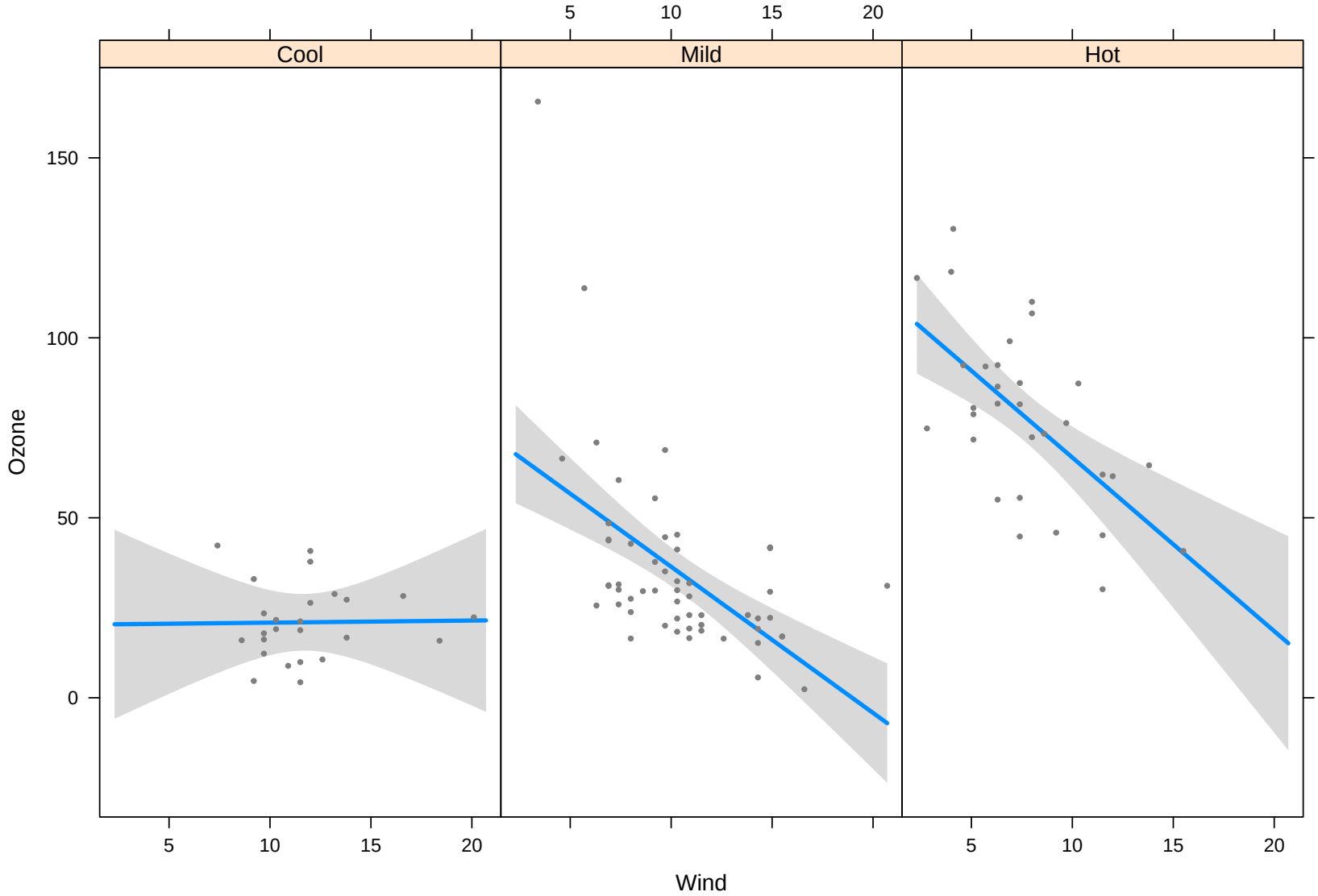
Das Paket visreg - Interaktionen

```
airquality$Heat <- cut(airquality$Temp, 3,
labels=c("Cool", "Mild", "Hot"))
fit <- lm(Ozone ~ Solar.R + Wind * Heat, data = airquality)
summary(fit)
```

```
##
## Call:
## lm(formula = Ozone ~ Solar.R + Wind * Heat, data = airqu
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.472 -11.640  -1.919   7.403 102.428
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.48042   17.38219   0.258 0.797102
## Solar.R         0.07634    0.02137   3.572 0.000538 ***
## Wind            0.05854    1.34860   0.043 0.965458
```

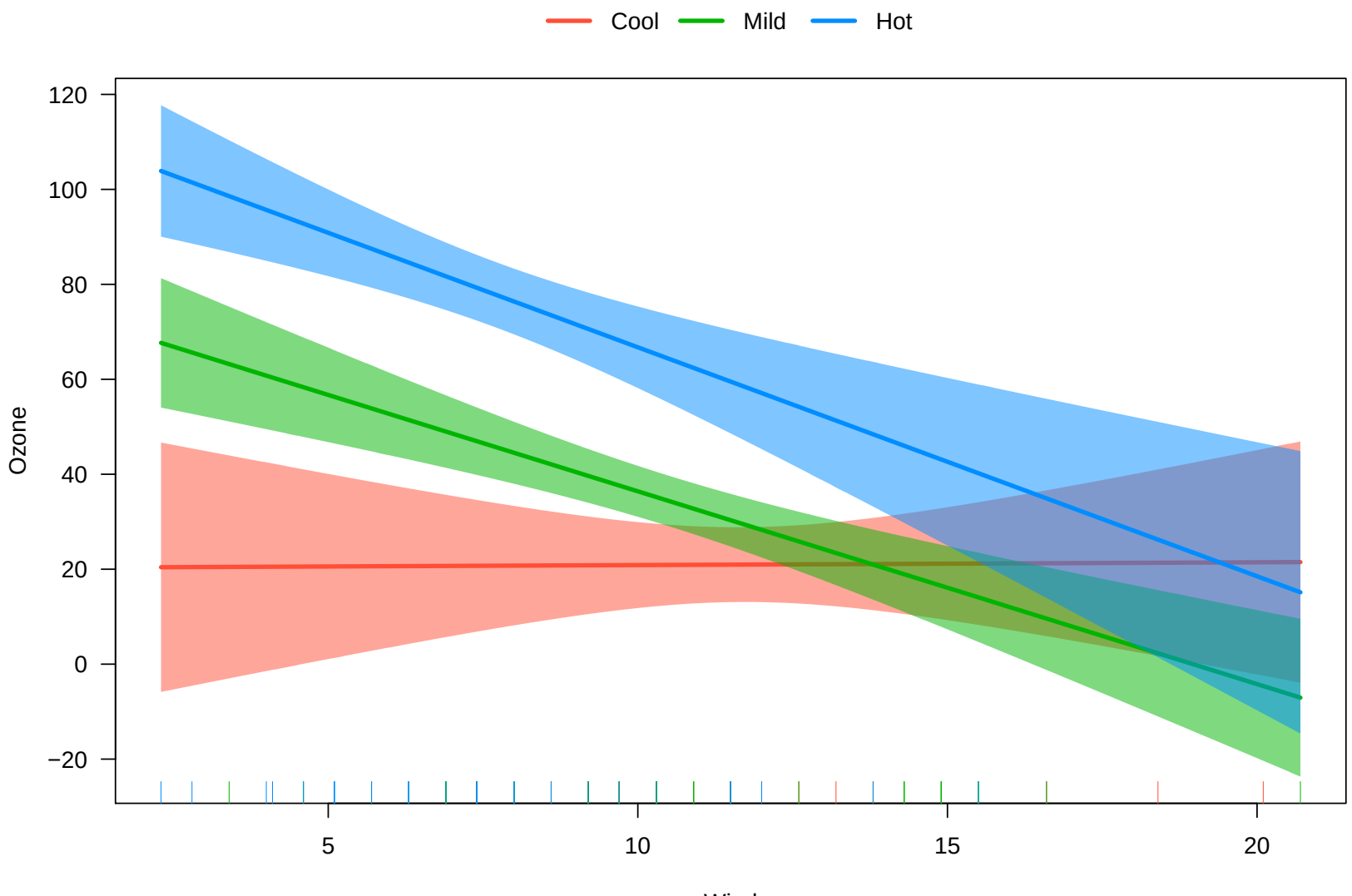
Steuern der Graphikausgabe mittels `layout`

```
visreg(fit, "Wind", by = "Heat",layout=c(3,1))
```



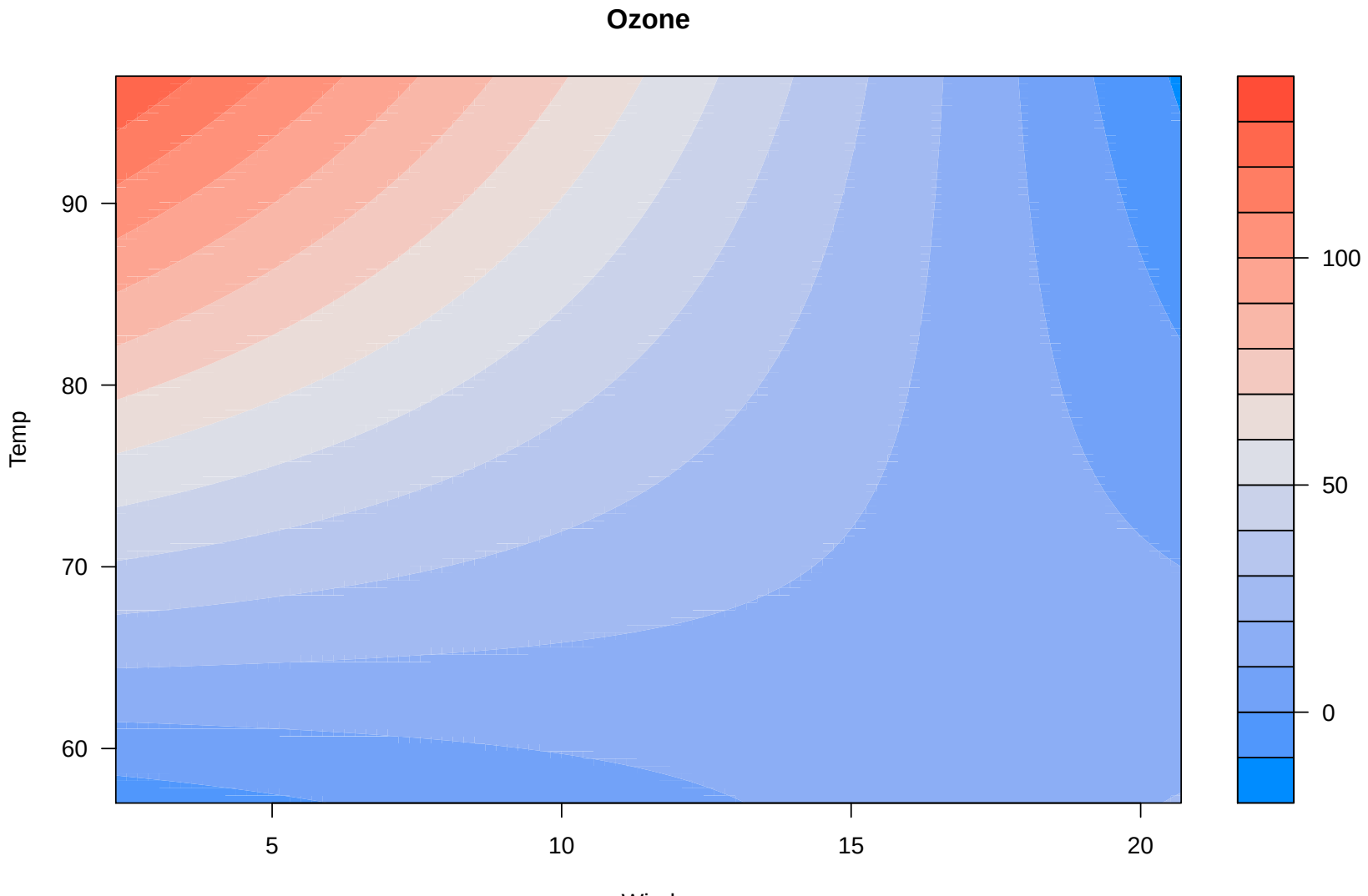
Das Paket visreg - Interaktionen overlay

```
fit <- lm(Ozone ~ Solar.R + Wind * Heat, data = airquality)
visreg(fit, "Wind", by="Heat", overlay=TRUE, partial=FALSE)
```



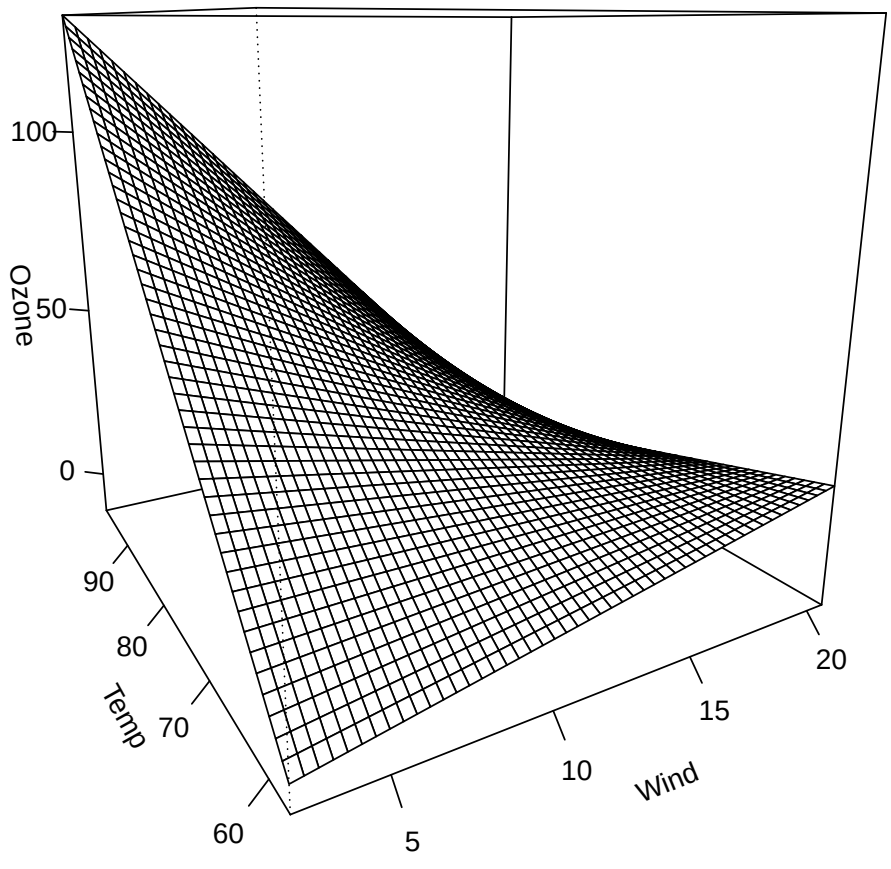
Das Paket visreg - visreg2d

```
fit2 <- lm(Ozone ~ Solar.R + Wind * Temp, data = airquality
visreg2d(fit2, "Wind", "Temp", plot.type = "image")
```



Das Paket visreg - surface

```
visreg2d(fit2, "Wind", "Temp", plot.type = "persp")
```



Linkliste - lineare Regression

- Regression - r-bloggers
- Das Komplette Buch von Faraway- sehr intuitiv geschrieben.
- Gute EinfÃ¼hrung auf Quick-R
- Multiple Regression
- Basis Regression - How to go about interpreting regression cofficients