

Aufgabe o4

Gruppe 01

13.5.2020

```
#Workspace laden
load("data/yingtan_20_ueb3.RData")
```

Aufgabe 10

a) Um den visuellen Eindruck aus Aufgabe 8b. mathematisch zu bestätigen, untersuchen Sie die Variable `Ca_exch` mit Hilfe eines gängigen Testverfahrens auf Normalverteilung.

```
#Test auf Normalverteilung mit dem Shapiro-Wilk-Test
shapiro.test(ljz$Ca_exch)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ljz$Ca_exch
## W = 0.90987, p-value = 2.914e-13
```

Der p-Wert ist $< 0,05$. Das bedeutet, dass eine Normalverteilung der Messwerte der austauschbaren Ca-Ionen unwahrscheinlich ist.

b) Führen Sie eine geeignete Transformation durch und plotten Sie das Dichte-Histogramm der transformierten Variablen. Wiederholen Sie den in 10a. verwendeten Test auf Normalverteilung und integrieren Sie die Testparameter ebenso in die Histogramm-Graphik wie die Kurve der zugehörigen Normalverteilung (curve und Normal). Wählen Sie eine gestrichelte Liniendarstellung doppelter Breite.

```
ljz_sqrt <- ljz
ljz_log10 <- ljz
ljz_log <- ljz

#Transformation
ljz_sqrt$Ca_exch <- sqrt(ljz_sqrt$Ca_exch)
ljz_log10$Ca_exch <- log10(ljz_log10$Ca_exch)
ljz_log$Ca_exch <- log(ljz_log$Ca_exch)
```

```
#Test auf Normalverteilung
shapiro.test(ljz_sqrt$Ca_exch)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ljz_sqrt$Ca_exch
## W = 0.98175, p-value = 0.0002949
```

```
shapiro.test(ljz_log10$Ca_exch)
```

```
##
```

```

## Shapiro-Wilk normality test
##
## data:  ljz_log10$Ca_exch
## W = 0.97684, p-value = 3.199e-05

shapiro.test(ljz_log$Ca_exch)

##
## Shapiro-Wilk normality test
##
## data:  ljz_log$Ca_exch
## W = 0.97684, p-value = 3.199e-05

#Histogram
myhist <- hist(ljz_sqrt$Ca_exch,
               breaks="FD", #Bin-Weite nach Freedman-Diaconis
               freq=FALSE, #Wahrscheinlichkeitsdichte statt Häufigkeit, alternative prob=TRUE
               col="grey",
               main="Dichte-Histogramm der transformierten austauschbaren Ca-Ionen",
               xlab="Transformierte austauschbare Ca-Ionen",
               ylab="Dichte",
               las=1)

#Dichte als Funktion
lines(x=density(ljz_sqrt$Ca_exch))

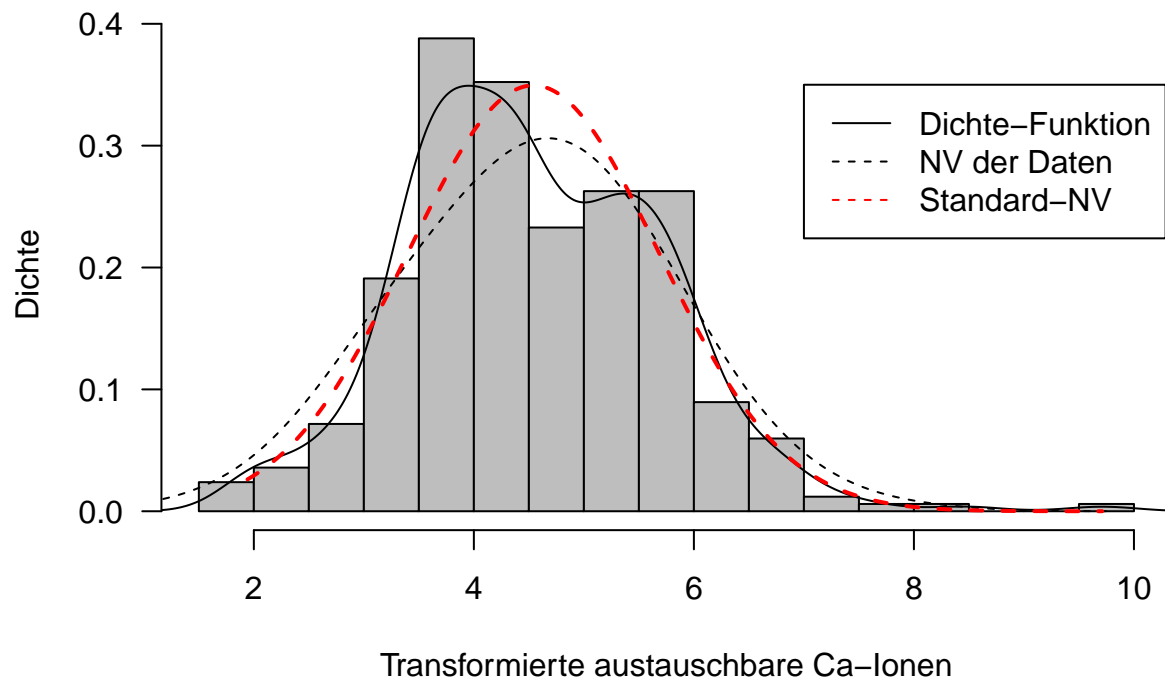
#Normalverteilung der Daten
norm=rnorm(length(ljz_sqrt$Ca_exch),
            mean = mean(ljz_sqrt$Ca_exch),
            sd= sd(ljz_sqrt$Ca_exch))
lines(x=density(norm, adjust = 2), lty = 2)

#Standard-Normalverteilung
xfit <- seq(min(ljz_sqrt$Ca_exch),
            max(ljz_sqrt$Ca_exch),
            length=length(ljz_sqrt$Ca_exch))
yfit <- dnorm(xfit,
              mean=mean(ljz_sqrt$Ca_exch),
              sd=sd(ljz_sqrt$Ca_exch))

lines(xfit, yfit, col="red", lwd= 2, lty=2)
legend(7, 0.35,
       legend = c("Dichte-Funktion", "NV der Daten", "Standard-NV"),
       col = c("black", "black", "red"),
       lty = c(1,2,2))

```

Dichte-Histogramm der transformierten austauschbaren Ca-Ionen



Der p-Wert ist nach wie vor $<0,05$, allerdings deutlich dichter dran als vor der Transformierung. Der W-Wert ist, wie auch schon vor der Transformierung, idalerweise nahe 1.

- c) Beurteilen Sie abschließend in knappen Worten, ob die Transformation erfolgreich war, oder ob sich keine entscheidende Annäherung an eine Normalverteilung herstellen ließ.

Aufgabe 11 Räumliche Daten in R

- a) Überführen Sie das Objekt `ljz` in eines der Klasse `SpatialPointsDataFrame`. Plotten Sie nun die Daten entsprechend ihrer geographischen Position. Beschriften Sie geschickt und fügen Sie ihrer Graphik ein rechtwinkliges Grid hinzu. Was lässt sich über das Sampling-Design sagen? Schätzen Sie die maximale Entfernung zweier Punkte im Untersuchungsgebiet ab.

```
library(sp)

#Spalten mit den Koordinaten selektieren
spatialCoords <- dplyr::select(ljz,
                                long = EAST,
                                lat = NORTH)

#Koordinatensystem definieren
coordRefSys <- CRS("+proj=utm +zone=50 +ellps=WGS84")

#SpatialPointsDataFrame erstellen
SPDFljz <- SpatialPointsDataFrame(spatialCoords,
                                   ljz,
                                   proj4string = coordRefSys)
```

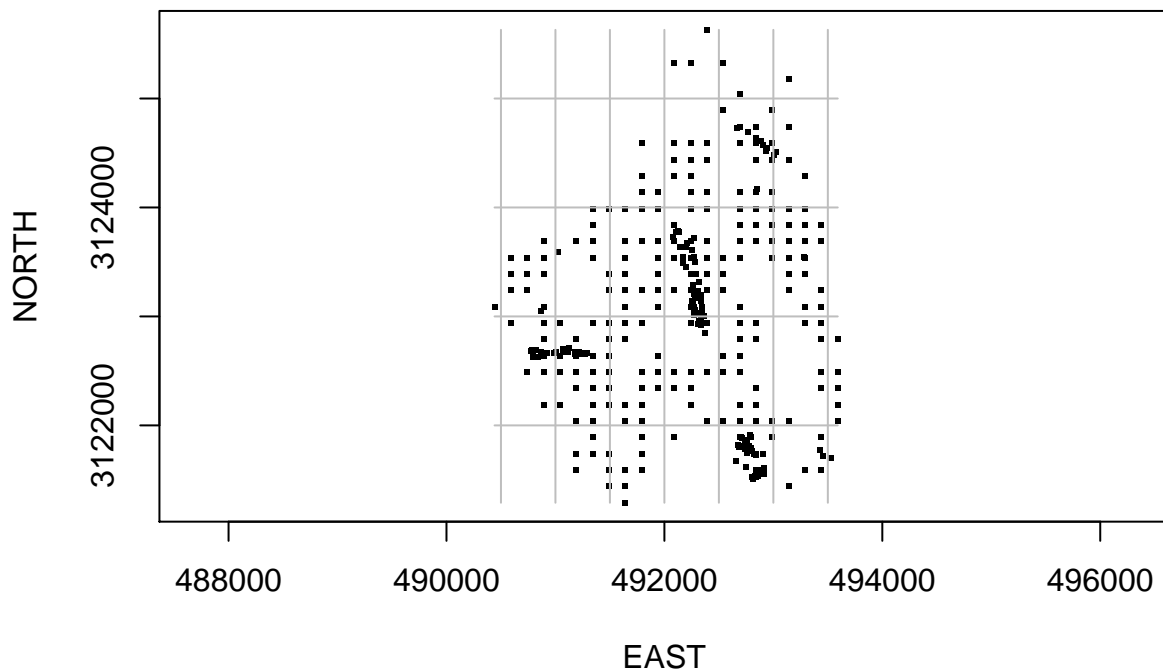
```

#graphische Darstellung
grid <- gridlines(SPDFljz,
                  easts = pretty(bbox(SPDFljz)[1,]),
                  norths = pretty(bbox(SPDFljz)[2,]))

plot(SPDFljz,
     main = "Orte der Probenentnahme in Yingtan",
     axes = TRUE,
     pch = ".",
     cex = 3,
     xlab = "EAST",
     ylab = "NORTH",
     xlim = c(490000, 494000))
plot(grid,
     col = "grey",
     add = TRUE)

```

Orte der Probenentnahme in Yingtan



```

#Distanzen abschaetzen
bbox(SPDFljz)

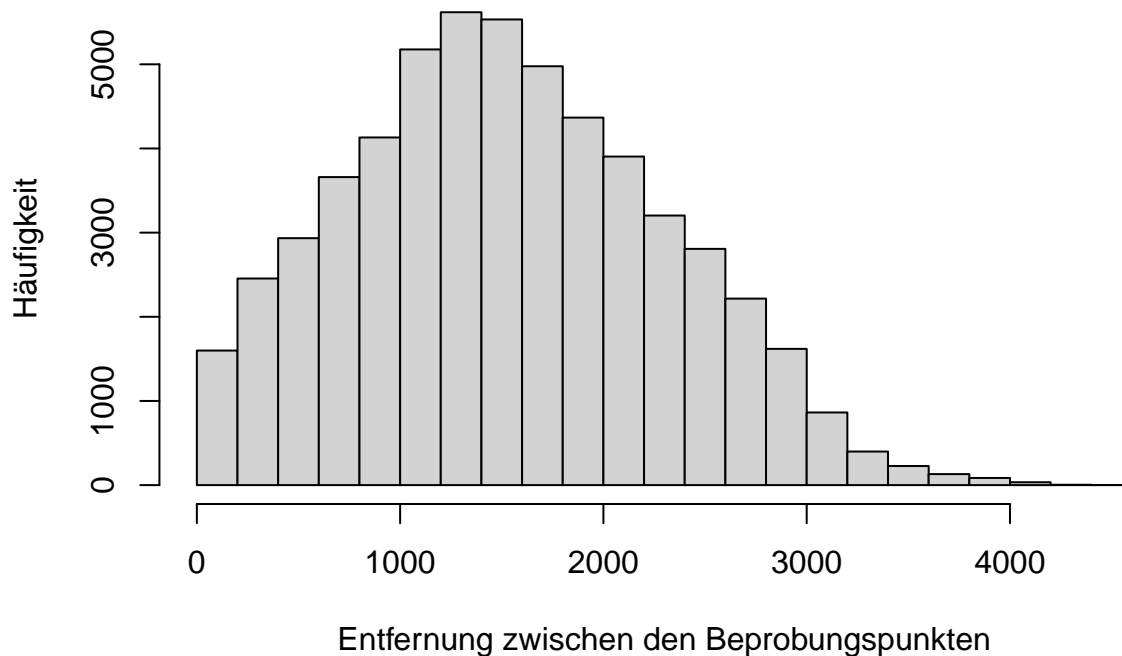
##          min      max
## long  490441  493591
## lat   3121290 3125630

ptsdist <- dist(coordinates(SPDFljz))
hist(ptsdist,
     main = "Histogramm der Abstände der Beprobungsstellen",

```

```
xlab = "Entfernung zwischen den Beprobungspunkten",
ylab = "Häufigkeit")
```

Histogramm der Abstände der Beprobungsstellen



```
max(ptsdist)
```

```
## [1] 4404.327
```

Größtenteils wurde die Fläche gleichmäßig beprobt. An einigen Stellen allerdings gar nicht und an anderen liegen die beprobten Stellen sehr nah bei einander. Das Untersuchungsgebiet liegt zwischen 490441 und 492591 für die Rechtswerte und zwischen 3121290 und 3125630 für die Hochwerte. Die maximale Distanz der beprobten Stellen beträgt etwa 4400m.

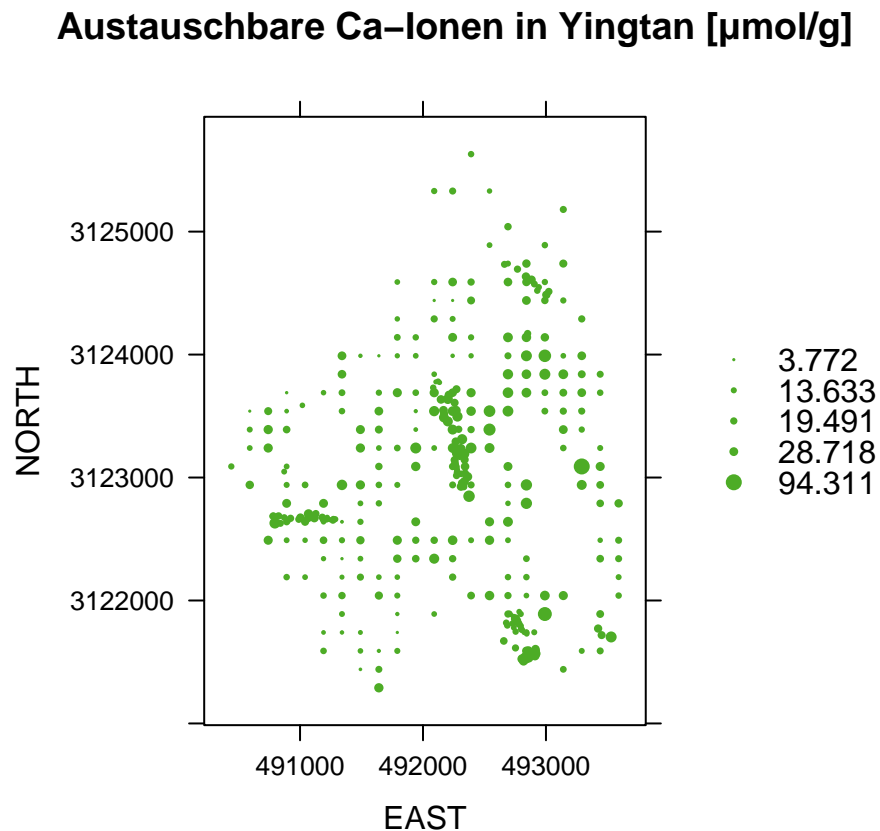
- b) Verschaffen Sie sich anschließend einen Überblick über die Lage der Punkte indem Sie diese vor einem Satellitenbild plotten (mapview). Lassen sich durch das Satellitenbild weitere Äußerungen über das Sampling-Design treffen? Hinweis: Um das richtige Satellitenbild zu finden, muss dem Objekt ljk das richtige Koordinatensystem zugewiesen werden. Denken Sie dazu an die Objektbeschreibung in Übung 2.

```
library(mapview)
print(mapview(SPDFljk, map.type = "Esri.WorldImagery"))
```

Das Beprobungsgebiet wird offensichtlich von einer Autobahn oder größeren Straße durchquert. Einige Standorte der Probenentnahmestellen liegen sehr nah an der Fahrbahn. Das kann die chemische Zusammensetzung beeinflussen und so das Ergebnis für das Untersuchungsgebiet beeinflussen. Lücken in den Beprobungsstellen sind durch Gewässer oder Siedlungen zu erklären. Neben dem regulären regelmäßigen Beprobungsmuster gibt es vier Bereiche, die entlang von Catenen intensiver beprobt wurden (s. Spalte SAMPLING). Hierbei ist nicht zu erkennen, ob diese speziellen "Areas-of-Interests" zufällig oder durch eine bestimmte Sampling-Strategie gewählt wurden.

- c) Erstellen Sie mit Hilfe der Methode `bubble` einen Plot, der Ihnen zusätzlich zur geographischen Position jeweils die Größenordnung des Ca-Gehalts (untransformiert) vermittelt. Setzen Sie die maximale Kreisstärke auf 1 und sorgen Sie dafür, dass die Koordinatenachsen geplottet werden. Gehen Sie der Frage nach, ob ein regionaler Trend auszumachen ist. Liegen die Extremwerte konzentriert vor, oder verteilen sie sich auf das gesamte Einzugsgebiet?

```
bubble(SPDFljz, "Ca_exch",  
  main = "Austauschbare Ca-Ionen in Yingtan [ $\mu\text{mol/g}$ ]",  
  xlab = "EAST",  
  ylab = "NORTH",  
  scales = list(draw = TRUE),  
  maxsize = 1)
```



Die Proben, die nah an Bahnleisen entnommen wurden, weisen eine höhere Konzentration an austauschbaren Ca-Ionen auf. Darüber hinaus ist sie im Osten, besonders östlich der zentralen Catena, des Beprobungsgebiets etwas höher als im Westen. Ein klarer Unterschied in Nord-Süd-Richtung ist nicht zu erkennen.