

Transmission Risk Comparison

Remo Schmutz

2022-12-21

Libraries

```
library(tidyverse)
library(ggplot2)
require("knitr")
library(gridExtra)
library(grid)
library(lubridate)
library(dplyr)
library(hms)
library(truncdist)
library(crch)
library(stats)
library(LaplacesDemon)
library(ggstatsplot)
library(MASS)
library(fitdistrplus)
library(truncnorm)
```

Data

```
ch <- readRDS("data-clean/co2-ch.rds") #swiss data
satz <- readRDS("data-clean/co2-sa-tz.rds")

ch <- ch %>%
  filter(co2 > 400) %>%
  arrange(co2)

ch1 <- ch %>%
  filter(school == "School 1")

ch2 <- ch %>%
  filter(school == "School 2")

sa <- satz %>%
  filter(country == "South Africa") %>%
  filter(co2 < 3000) %>%
  filter(co2 > 400) %>%
```

```

  arrange(co2) #south africa data

tz <- satz %>%
  filter(country == "Tanzania") %>%
  filter(co2 < 3000) %>%
  filter(co2 > 400) %>%
  mutate(time_h = hour(date)) %>%
  filter(time_h >= 8) %>%
  arrange(co2) #tanzania data

```

Methods

Indoor Co2 concentration

* mean or distribution from data

* C_o := Outdoor Co2 concentration * from literature https://www.fsis.usda.gov/sites/default/files/media_file/2020-08/Carbon-Dioxide.pdf

* C_a := Volume fraction of CO2 added to exhaled breath during breathing

* Persily and de Jonge [Table 3 and 4] doi: 10.1111/ina.12383

* $\bar{f} := \int_{t=0}^{t=max} f dt$

* integrating over f values from different times (2) or using a distribution based on the data

* I := Number of infectors in the class

* estimated using prevalence of the age group in the country

* q := Quantum per hour

* assuming a distribution from literature

* t := time

* changing this parameter to compare

* n := number of people in the class

* data (Switzerland) or assumption (South Africa, Tanzania)

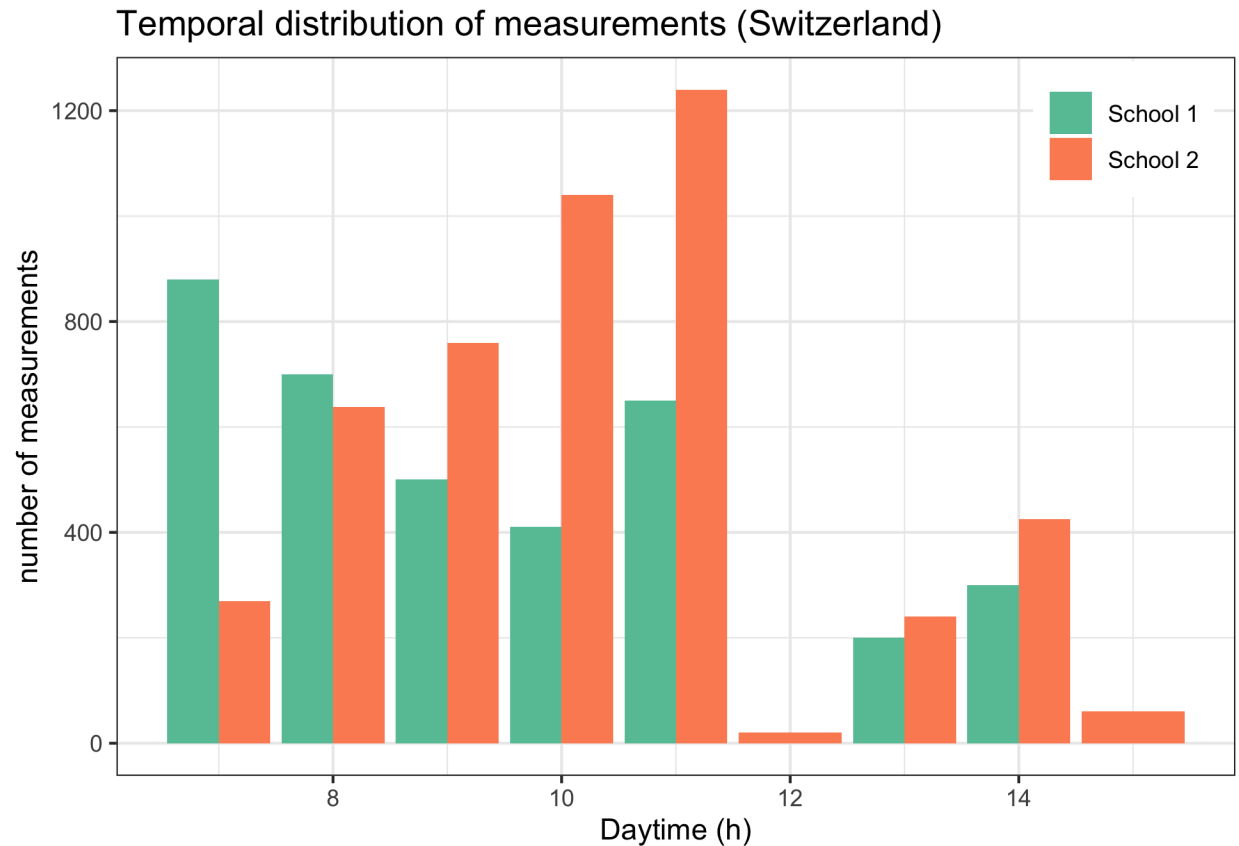
Preprocess

```

ch_hourly <- ch %>%
  mutate(time_h = hour(time)) %>%
  group_by(school, time_h) %>%
  summarize(mean = mean(co2),
             lower = quantile(co2, 0.25),
             upper = quantile(co2, 0.75),
             n_data = n()) %>%
  ungroup()

ch_hourly %>%
  ggplot(aes(x = time_h, y = n_data, fill = school)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  xlab("Daytime (h)") +
  ylab("number of measurements") +
  ggtitle("Temporal distribution of measurements (Switzerland)") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank()) +
  scale_color_brewer(palette = "Set2") +
  scale_fill_brewer(palette = "Set2")

```

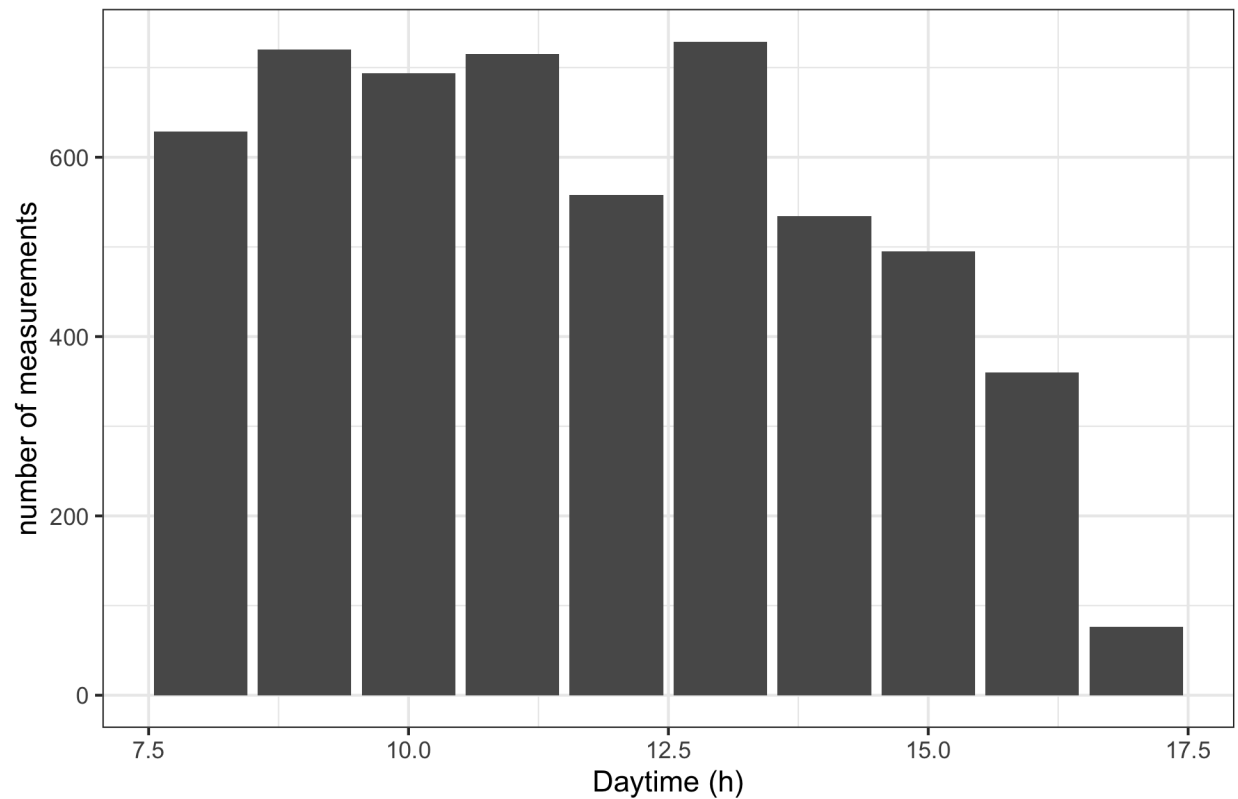


there is a reasonable number of data points per hour but missing data at hour 12 for school 2 (no les

```
tz_hourly <- tz %>%
  mutate(time_h = hour(date)) %>%
  group_by(time_h) %>%
  summarize(mean = mean(co2),
             lower = quantile(co2, 0.25),
             upper = quantile(co2, 0.75),
             n_data = n()) %>%
  ungroup()

tz_hourly %>%
  ggplot(aes(x = time_h, y = n_data)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  xlab("Daytime (h)") +
  ylab("number of measurements") +
  ggtitle("Temporal distribution of measurements (Tanzania)") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank())
```

Temporal distribution of measurements (Tanzania)



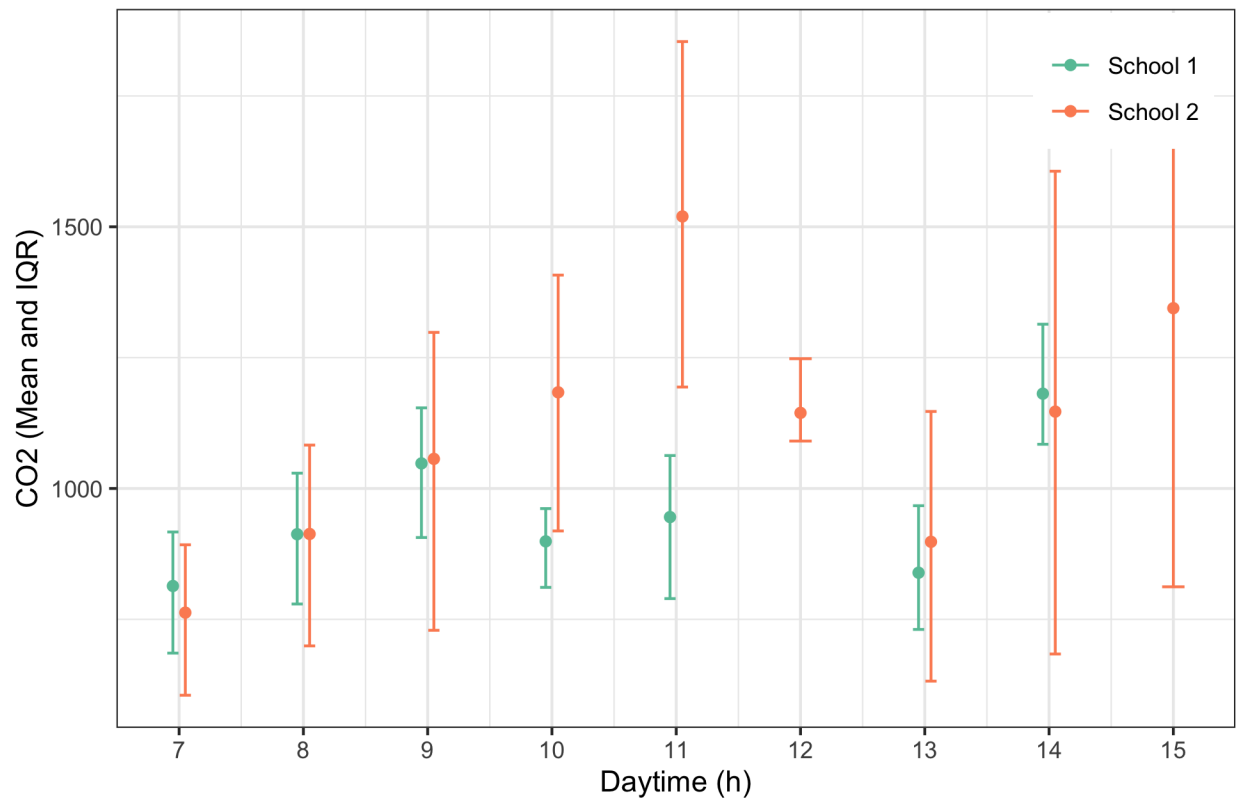
data is measured throughout the day in south africa

Analysis

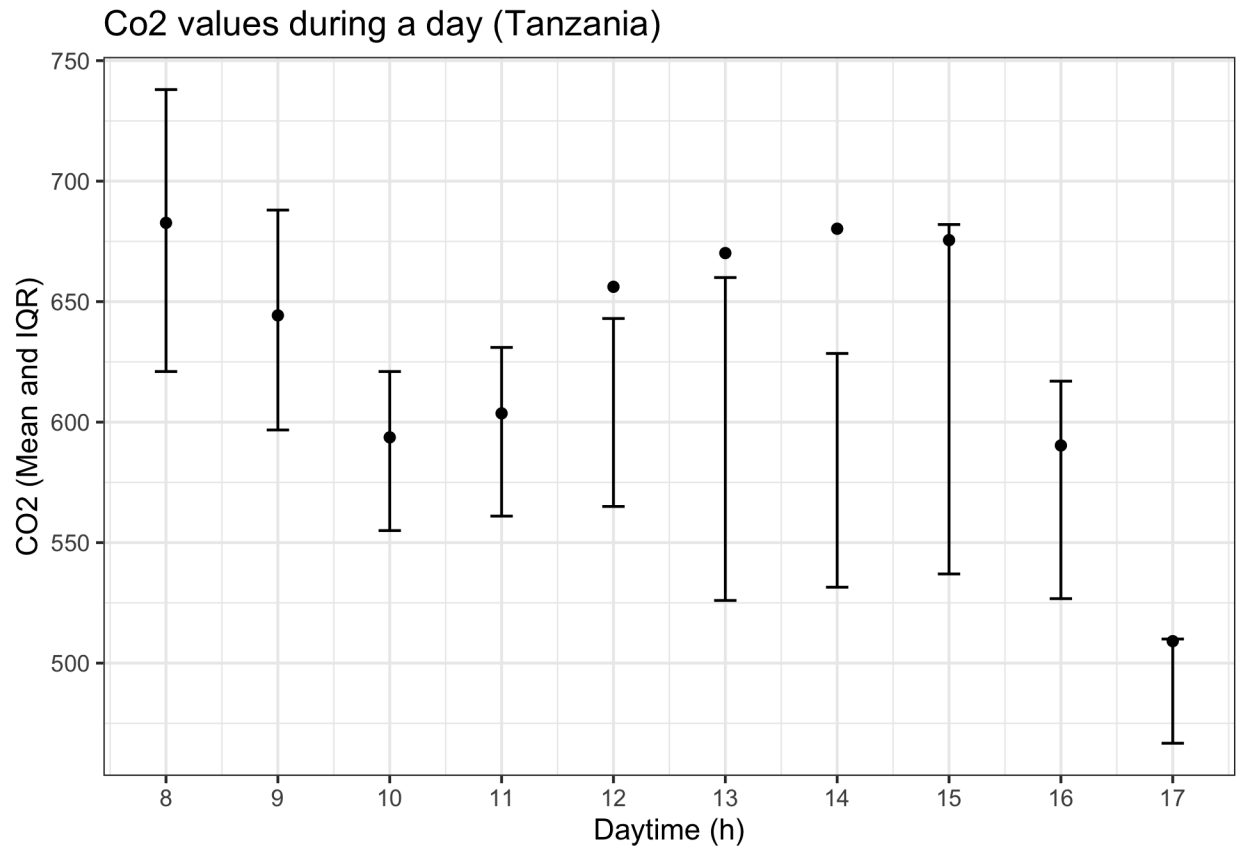
Co2 over time

```
ch_hourly %>% #plot co2 during the day (ch)
  ggplot(aes(x = time_h, group = school, color = school)) +
  geom_errorbar(aes(ymin = lower, ymax = upper), width = .2, position = position_dodge2(width = .2)) +
  geom_point(aes(y = mean), position = position_dodge2(width = .2)) +
  scale_color_brewer(palette = "Set2") +
  scale_x_continuous(breaks = seq(7, 16, 1)) +
  labs(x = "Daytime (h)", y = "CO2 (Mean and IQR)") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank()) +
  ggtitle("Co2 values during a day (Switzerland)")
```

Co2 values during a day (Switzerland)



```
tz_hourly %>% #plot co2 during the day (tz)
  ggplot(aes(x = time_h)) +
  geom_errorbar(aes(ymin = lower, ymax = upper), width = .2, position = position_dodge2(width = .2)) +
  geom_point(aes(y = mean), position = position_dodge2(width = .2)) +
  scale_color_brewer(palette = "Set2") +
  scale_x_continuous(breaks = seq(7, 17, 1)) +
  labs(x = "Daytime (h)", y = "CO2 (Mean and IQR)") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank()) +
  ggtitle("Co2 values during a day (Tanzania)")
```



#no time data available for south africa

Smoothing Data

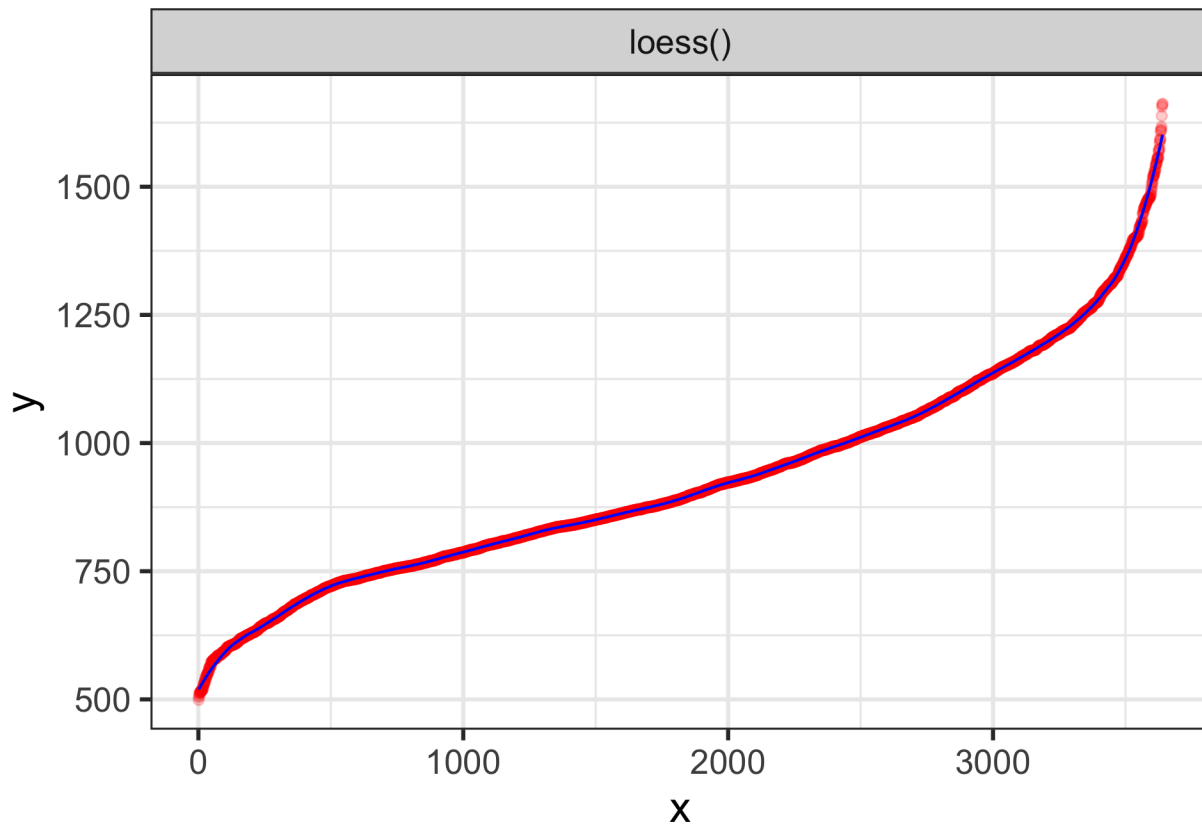
```
# CH
##school 1
set.seed(1)

dat_ch1 <- data.frame(
  x = 1:length(ch1$co2),
  y = ch1$co2
)

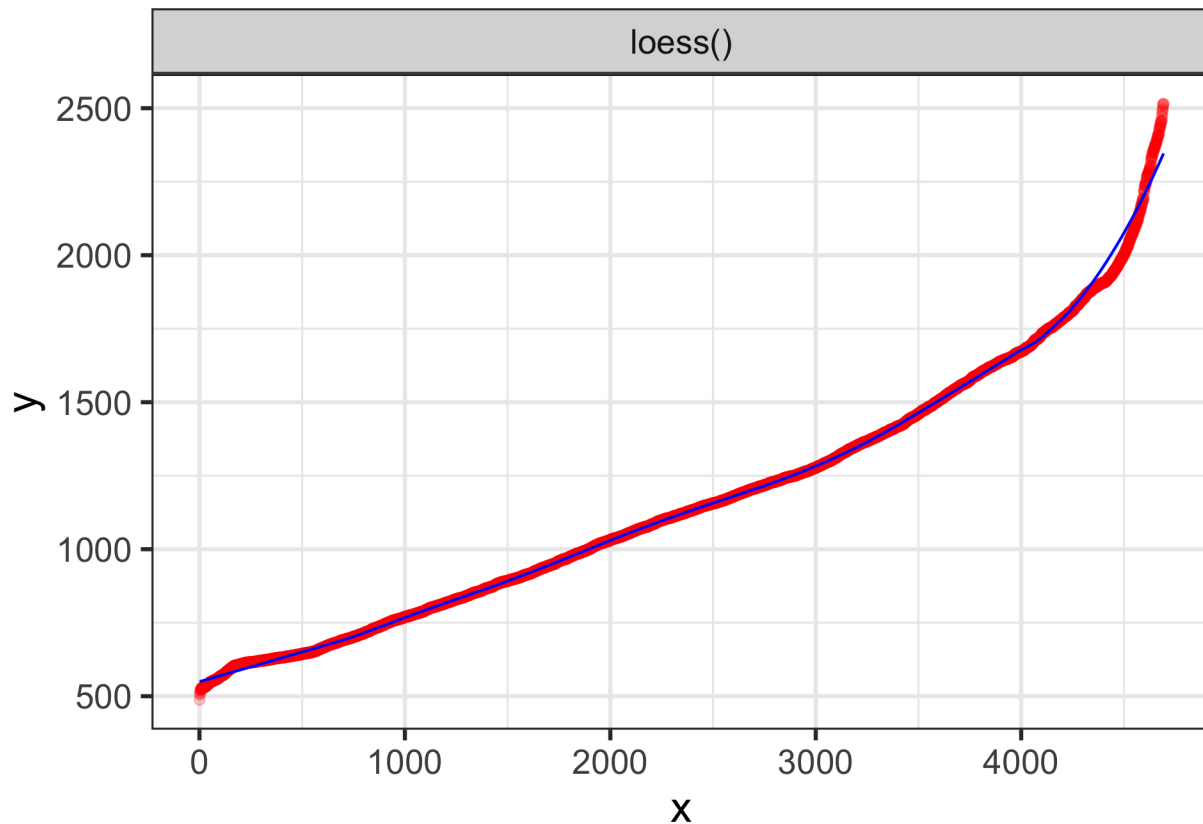
loessData_ch1 <- data.frame(
  x = 1:length(ch1$co2),
  y = predict(loess(y~x, dat_ch1, span = 0.1)),
  method = "loess()"
) %>%
  mutate(school = "school 1")

ggplot(loessData_ch1, aes(x, y)) +
  geom_point(dat = dat_ch1, aes(x, y), alpha = 0.2, col = "red") +
  geom_line(col = "blue") +
```

```
facet_wrap(~method) +  
theme_bw(16)
```



```
##school 2  
dat_ch2 <- data.frame(  
  x = 1:length(ch2$co2),  
  y = ch2$co2  
)  
  
loessData_ch2 <- data.frame(  
  x = 1:length(ch2$co2),  
  y = predict(loess(y~x, dat_ch2, span = 0.3)),  
  method = "loess()  
) %>%  
  mutate(school = "school 2")  
  
ggplot(loessData_ch2, aes(x, y)) +  
  geom_point(dat = dat_ch2, aes(x, y), alpha = 0.2, col = "red") +  
  geom_line(col = "blue") +  
  facet_wrap(~method) +  
  theme_bw(16)
```

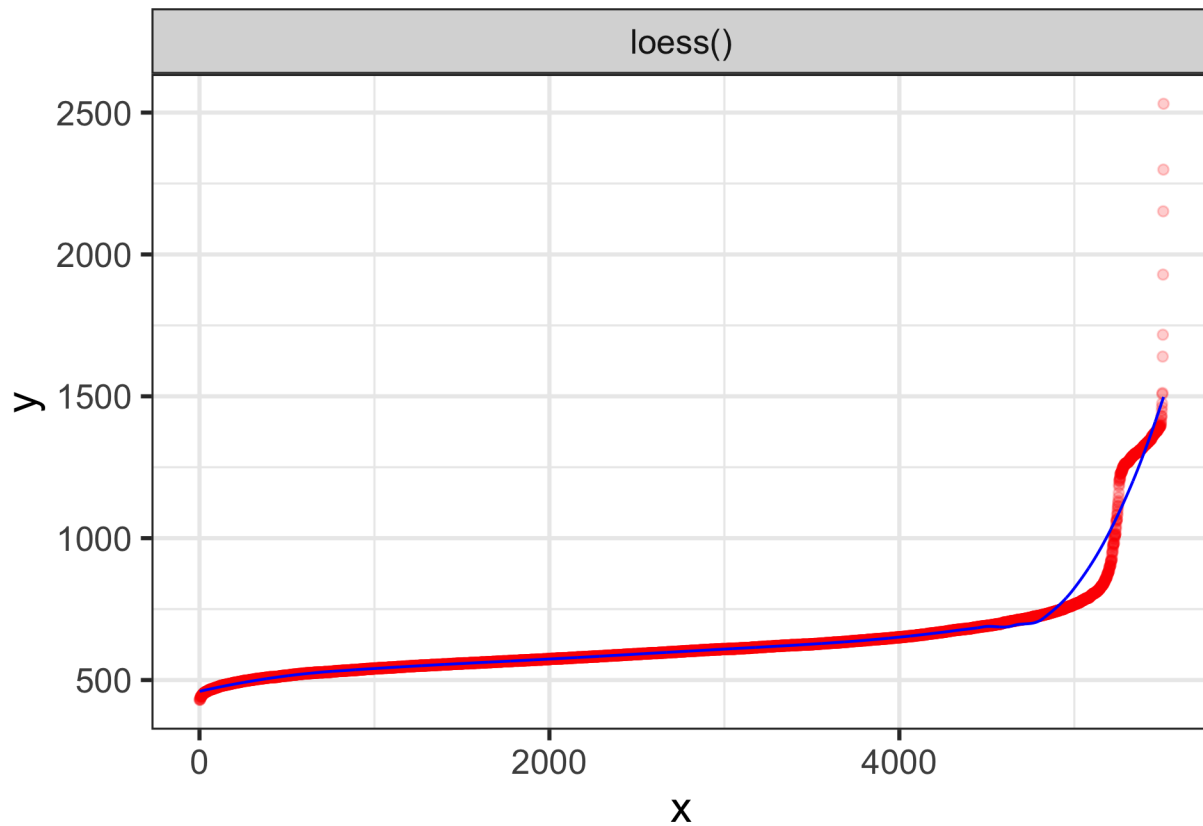


```
loessData_ch <- rbind(loessData_ch1,loessData_ch2)

# TZ
dat_tz <- data.frame(
  x = 1:length(tz$co2),
  y = tz$co2
)

loessData_tz <- data.frame(
  x = 1:length(tz$co2),
  y = predict(loess(y~x, dat_tz, span = 0.3)),
  method = "loess()"
)

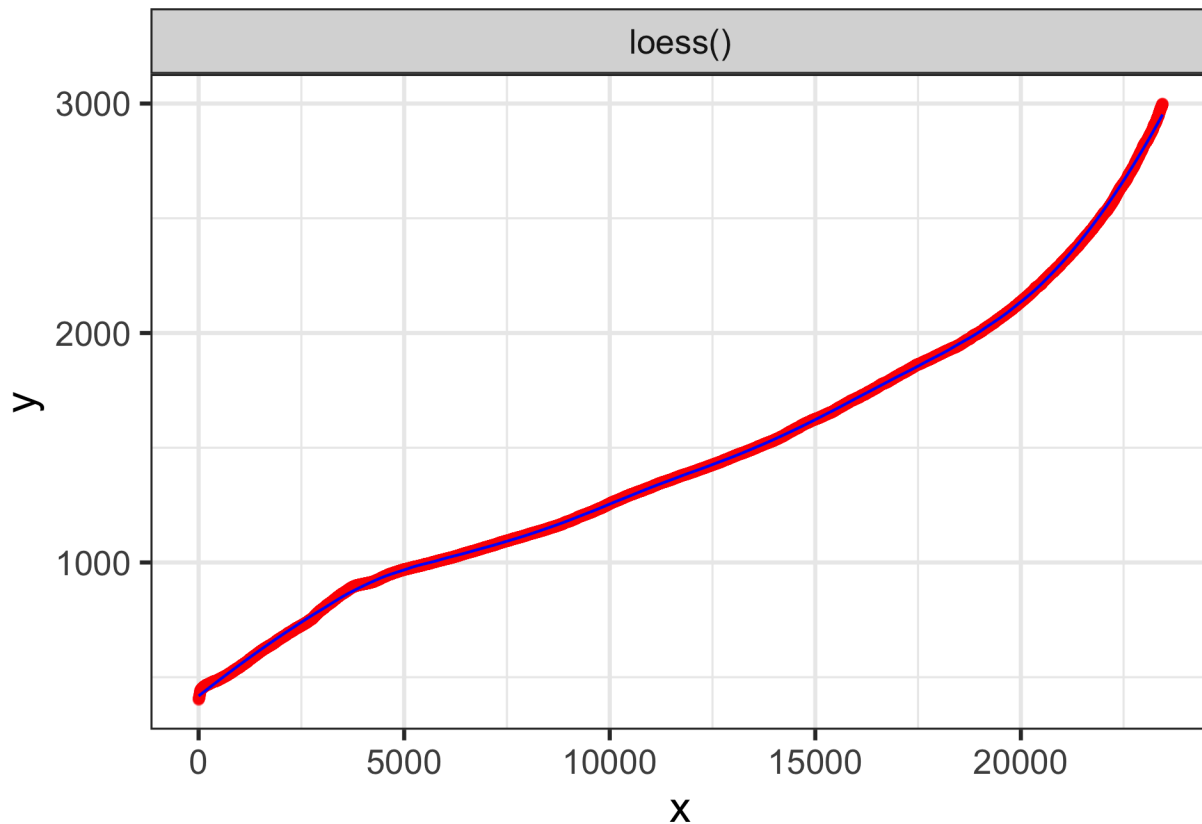
ggplot(loessData_tz, aes(x, y)) +
  geom_point(dat = dat_tz, aes(x, y), alpha = 0.2, col = "red") +
  geom_line(col = "blue") +
  facet_wrap(~method) +
  theme_bw(16)
```

```
# SA
dat_sa <- data.frame(
  x = 1:length(sa$co2),
  y = sa$co2
)

loessData_sa <- data.frame(
  x = 1:length(sa$co2),
  y = predict(loess(y~x, dat_sa, span = 0.3)),
  method = "loess()"
)

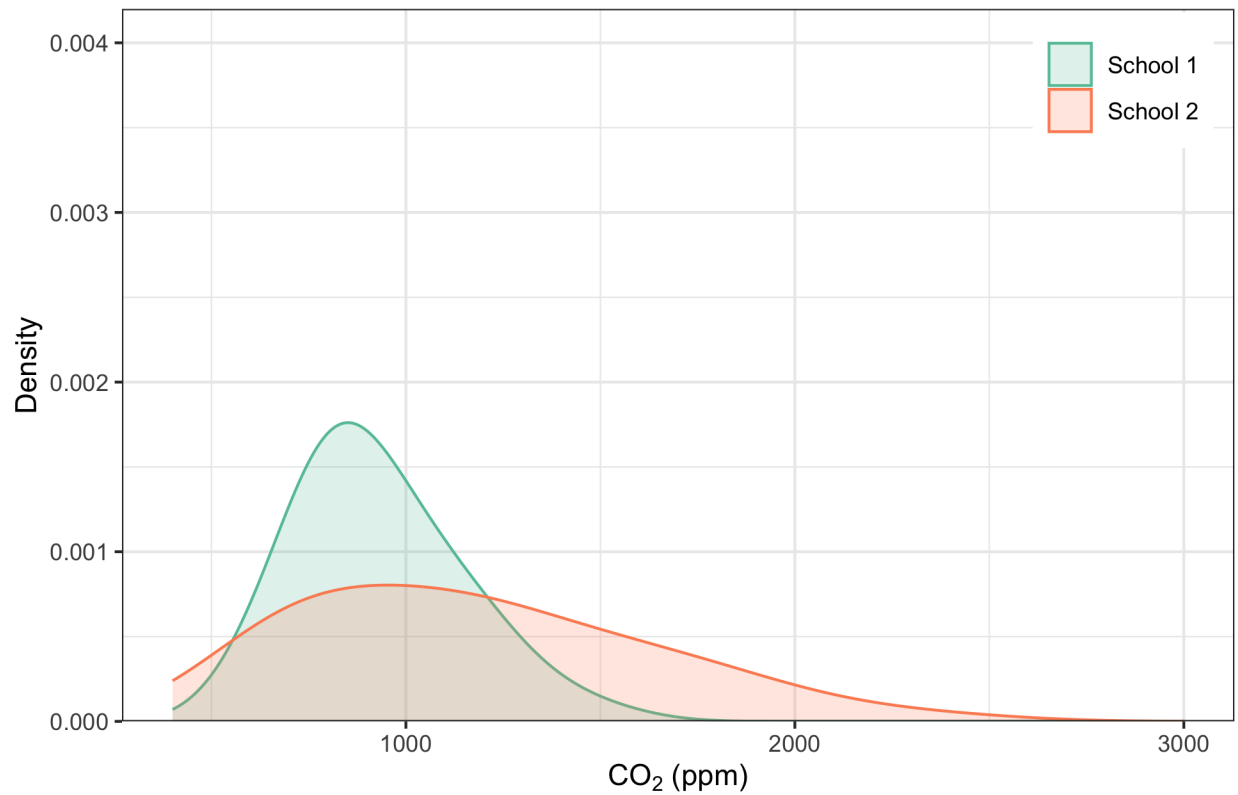
ggplot(loessData_sa, aes(x, y)) +
  geom_point(dat = dat_sa, aes(x, y), alpha = 0.2, col = "red") +
  geom_line(col = "blue") +
  facet_wrap(~method) +
  theme_bw(16)
```



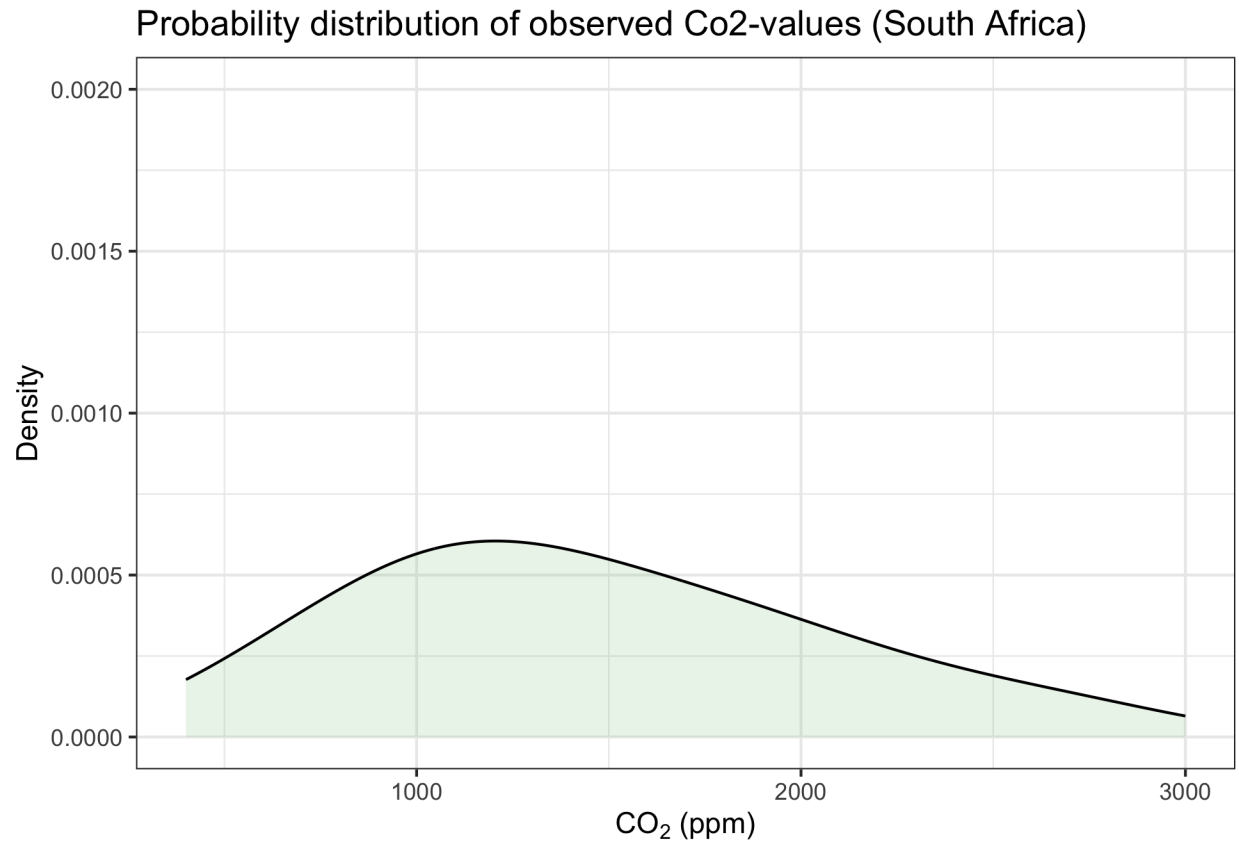
Co2 distribution

```
ch %>% #density plot ch
  ggplot(aes(x = co2, color = school, fill = school)) +
  geom_density(alpha = .2, kernel = "gaussian", adjust = 3) +
  scale_x_continuous(expand = c(0,0)) +
  scale_color_brewer(palette = "Set2") +
  scale_fill_brewer(palette = "Set2") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.05)), limits = c(0, 0.004)) +
  scale_x_continuous(limits = c(400, 3000)) +
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank()) +
  ggtitle("Probability distribution of observed Co2-values (Switzerland)")
```

Probability distribution of observed Co2-values (Switzerland)

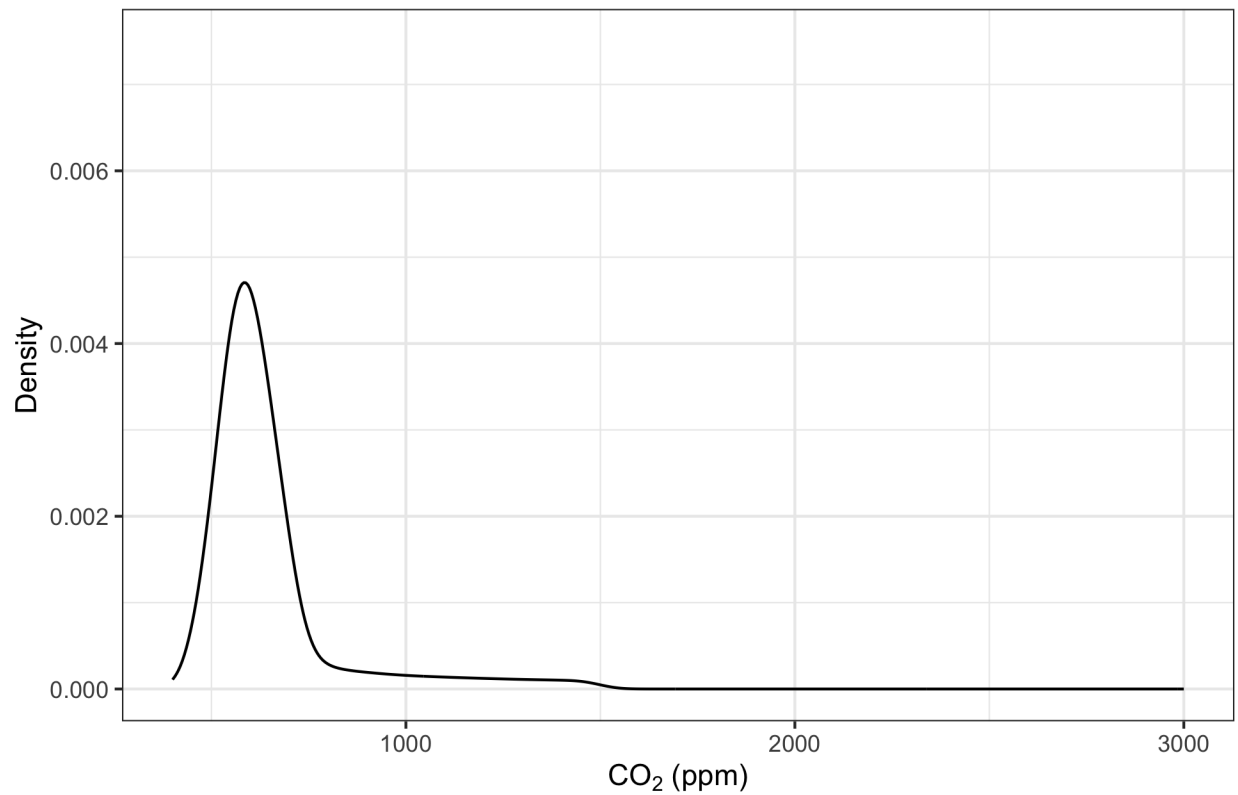


```
loessData_sa %>% #density plot sa smooth
  ggplot(aes(x = y)) +
  geom_density(alpha = .2, kernel = "gaussian", adjust = 4, fill = "darkseagreen3") +
  scale_x_continuous(expand = c(0,0)) +
  scale_color_brewer(palette = "Set2") +
  scale_fill_brewer(palette = "Set2") +
  scale_y_continuous(limits = c(0, 0.002)) +
  scale_x_continuous(limits = c(400, 3000)) +
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  theme(legend.position = "none") +
  theme_bw() +
  ggtitle("Probability distribution of observed Co2-values (South Africa)")
```



```
loessData_tz %>% #density plot tz smooth
  ggplot(aes(x = y)) +
  geom_density(alpha = .2, kernel = "gaussian", adjust = 3.2) +
  scale_x_continuous(expand = c(0,0)) +
  scale_color_brewer(palette = "Set2") +
  scale_fill_brewer(palette = "Set2") +
  scale_y_continuous(limits = c(0, 0.0075)) +
  scale_x_continuous(limits = c(400, 3000)) +
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank()) +
  ggtitle("Probability distribution of observed Co2-values (Tanzania)")
```

Probability distribution of observed Co2-values (Tanzania)



Distribution co2

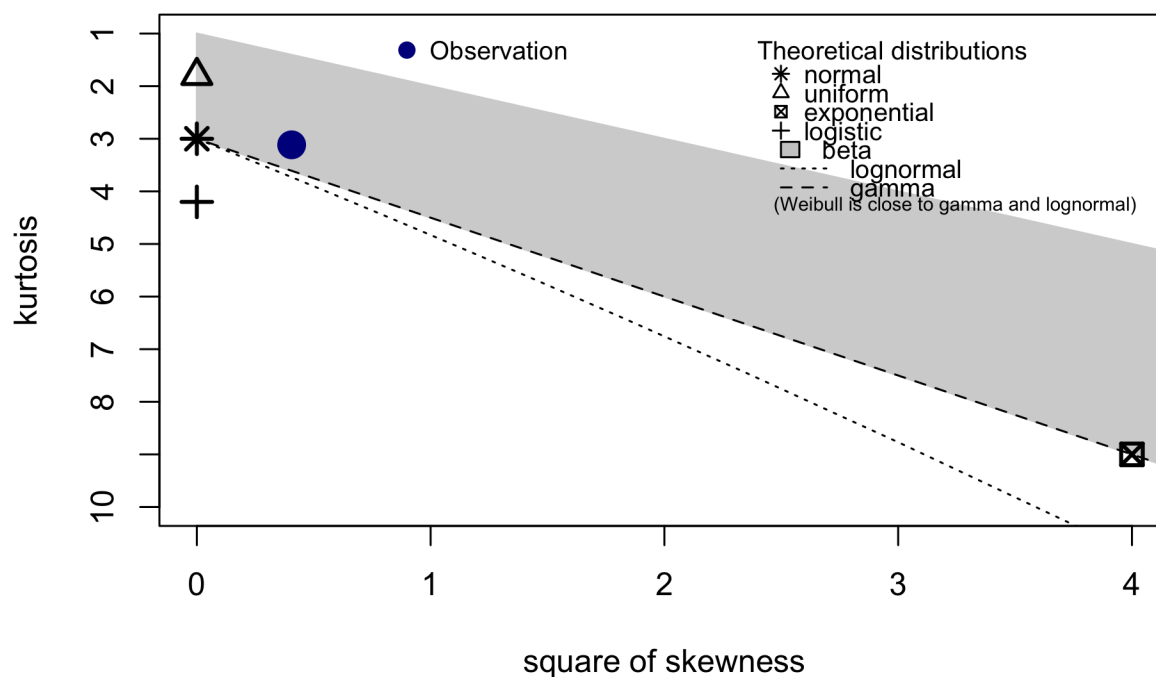
```
#C_a
C_a <- (((0.0048 + 0.0041 + 0.0053 + 0.0042)/4)*60)/8 #https://www.ncbi.nlm.nih.gov/pmc/articles/PMC566

#C_o
C_o <- 400 #p.p.m (taking a higher estimate because higher values ar possible when a lot of traffic ect
## all schools are directly on the side of a road (no info for tanzania), so i won't make a distinction

#school1
x_ch1 <- ch %>%
  filter(school == "School 1") %>%
  pull(co2)

descdist(x_ch1, discrete = FALSE) #normal distribution fits well
```

Cullen and Frey graph



```
## summary statistics
## -----
## min: 499 max: 1661.83
## median: 890.79
## mean: 929.6862
## estimated sd: 211.422
## estimated skewness: 0.6368133
## estimated kurtosis: 3.115236
```

```
fitdistr(x_ch1, "gamma") #get parameters
```

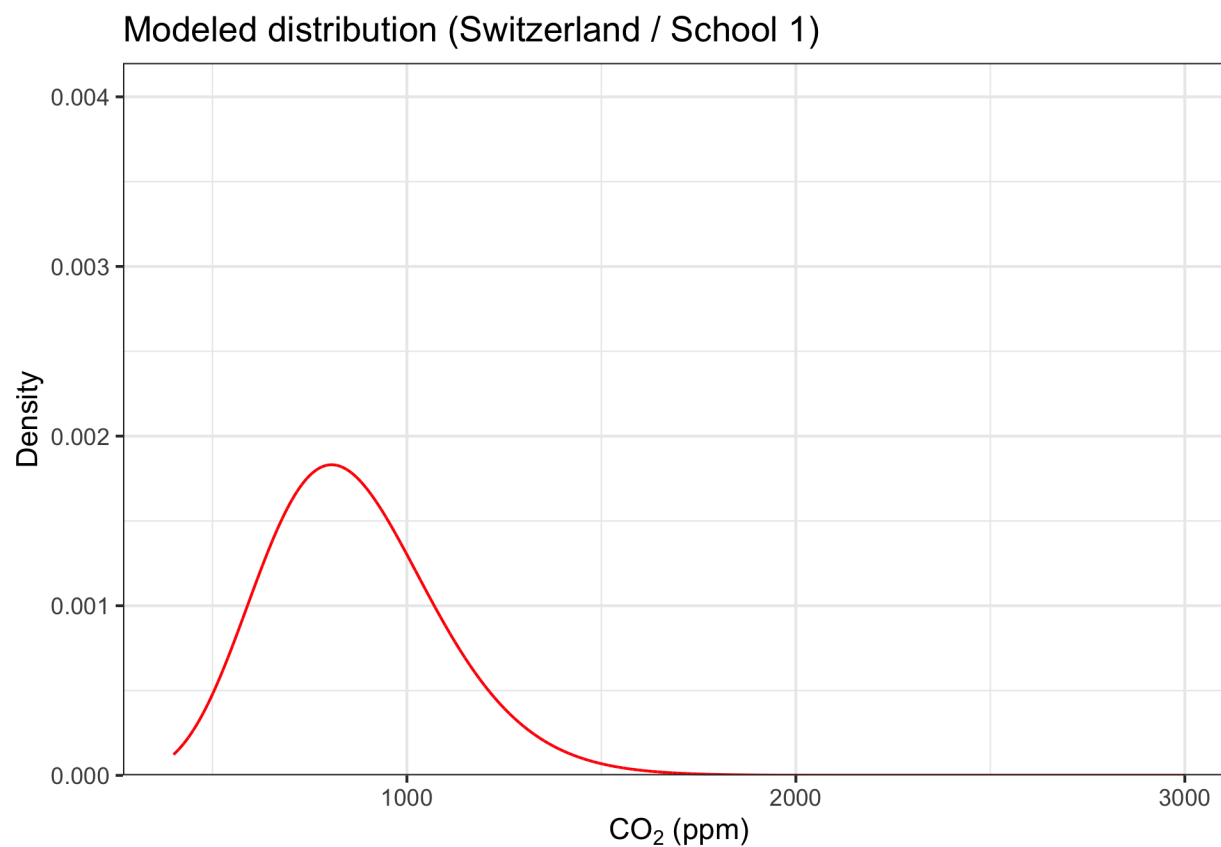
```
##          shape          rate
## 20.0718896059 0.0215899649
## ( 0.4539820835) ( 0.0004939311)
```

```
x <- seq(400, 3000, by = .1)
```

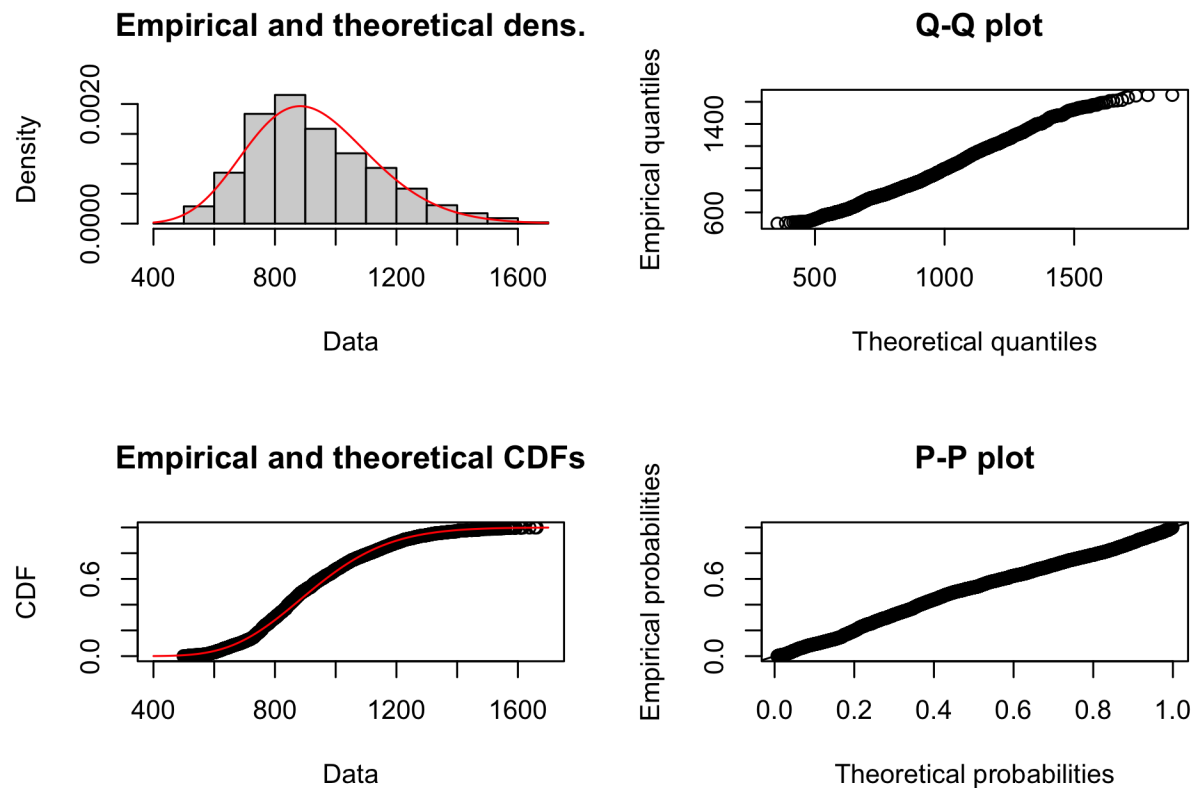
```
y_ch1 <- dtrunc(x, spec = "gamma", a = 400, b = 3000, shape = 14.71, rate = 0.017)
x_ch1_gamma <- data.frame(cbind(x, y_ch1))
```

```
x_ch1_gamma %>%
  ggplot(aes(x=x, y=y_ch1)) +
  geom_line(color = "red") +
  ggtitle("Modeled distribution (Switzerland / School 1)") + #plot density
```

```
labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
theme_bw() +
scale_y_continuous(expand = expansion(mult = c(0, 0.05)), limits = c(0, 0.004)) +
scale_x_continuous(limits = c(400, 3000))
```



```
ch1_fit_gamma <- fitdist(x_ch1, "gamma", lower=c(0,0)) #different fitting function
plot(ch1_fit_gamma) #plots comparison
```



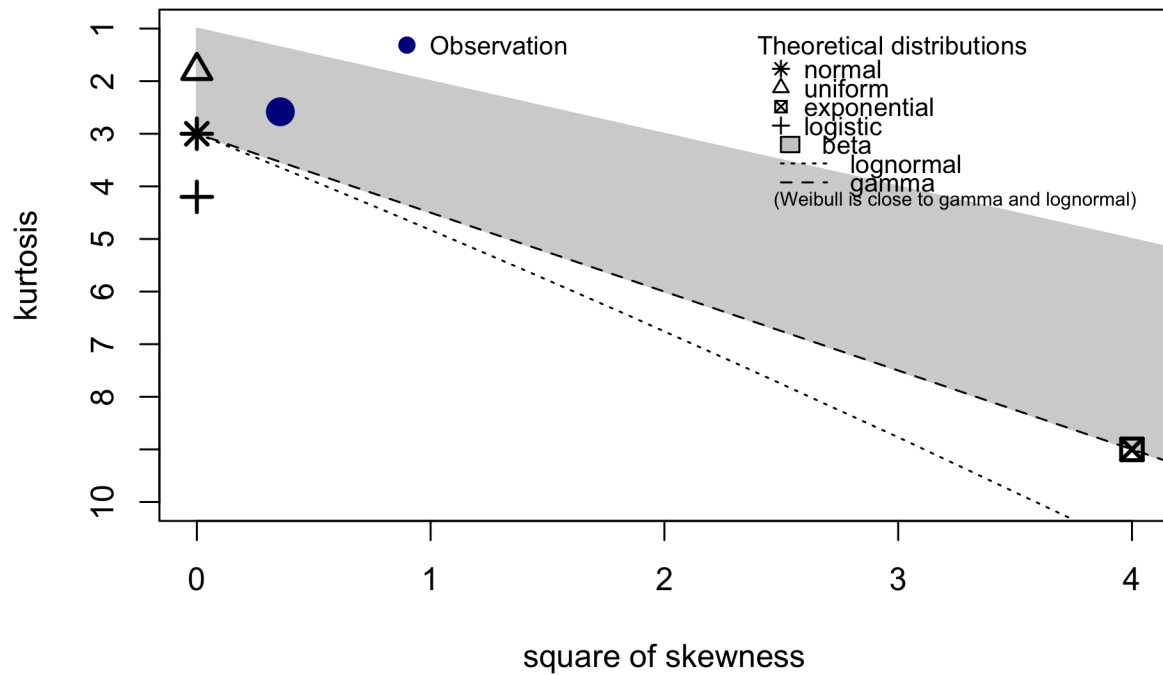
```
co2_distr_ch1 <- data.frame(co2 = seq(400, 3000, .1)) %>%
  mutate(prob = dtrunc(co2, spec = "gamma", a = 400, b = 3000, shape = 14.71, rate = 0.017) )

sample_co2_ch1 <- sample(co2_distr_ch1$co2, 10000, replace = TRUE, prob = co2_distr_ch1$prob) #sample c
sample_f_ch1 <- tibble(co2 = sample_co2_ch1, f = ((co2-C_o)/C_a)/1000000) %>% #sample f
  dplyr::select(-co2)
```

```
#school 2
x_ch2 <- loessData_ch2 %>%
  pull(y)

descdist(x_ch2, discrete = FALSE) #gamma fits ok
```


Cullen and Frey graph



```
## summary statistics
## -----
## min: 549.5165 max: 2346.033
## median: 1120.103
## mean: 1175.967
## estimated sd: 438.1166
## estimated skewness: 0.5976435
## estimated kurtosis: 2.581763
```

```
fitdistr(x_ch2, "normal") #get parameters
```

```
##      mean      sd
## 1175.967154 438.069961
## ( 6.394671) ( 4.521715)
```

```
fitdistr(x_ch2, "gamma")
```

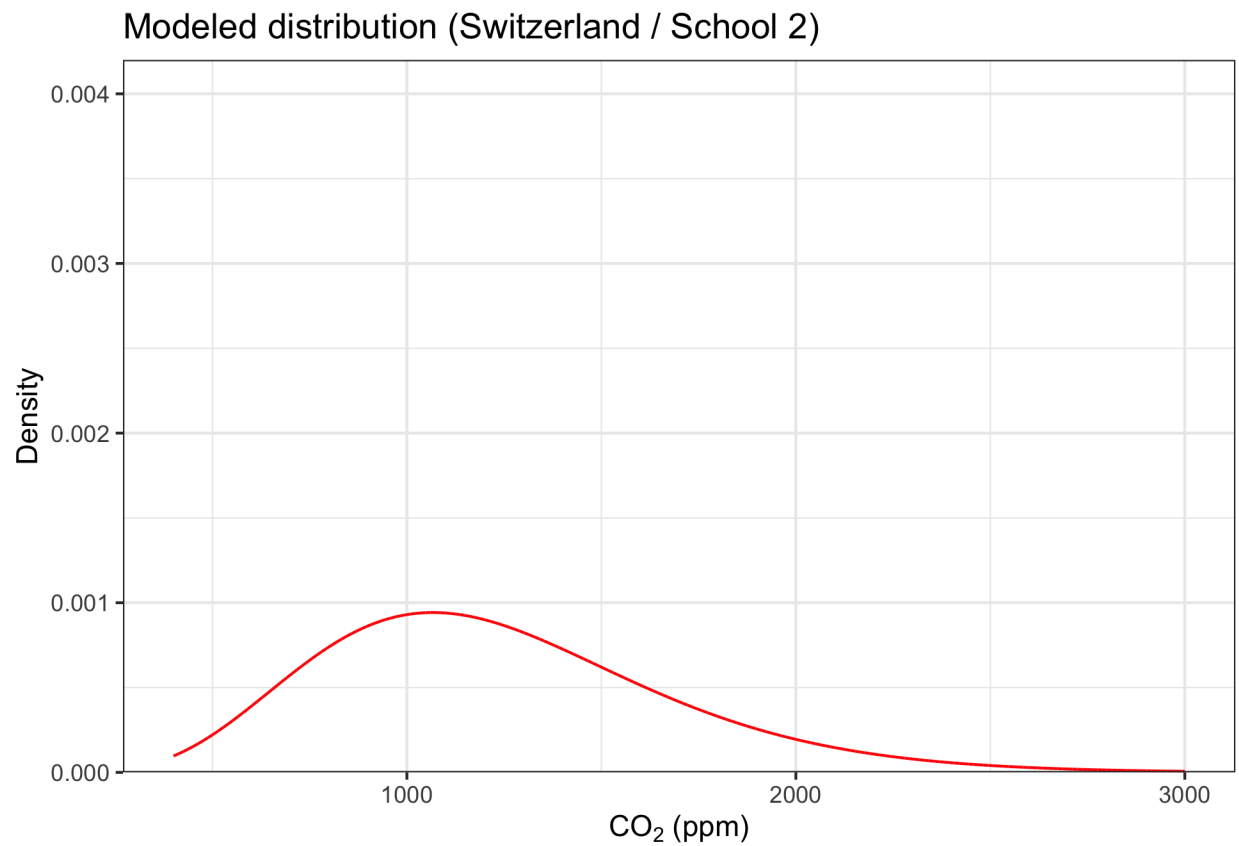
```
##      shape      rate
## 7.3808794503 0.0062764347
## (0.1338510866) (0.0001162775)
```

```
x <- seq(400, 3000, by = .1)
```

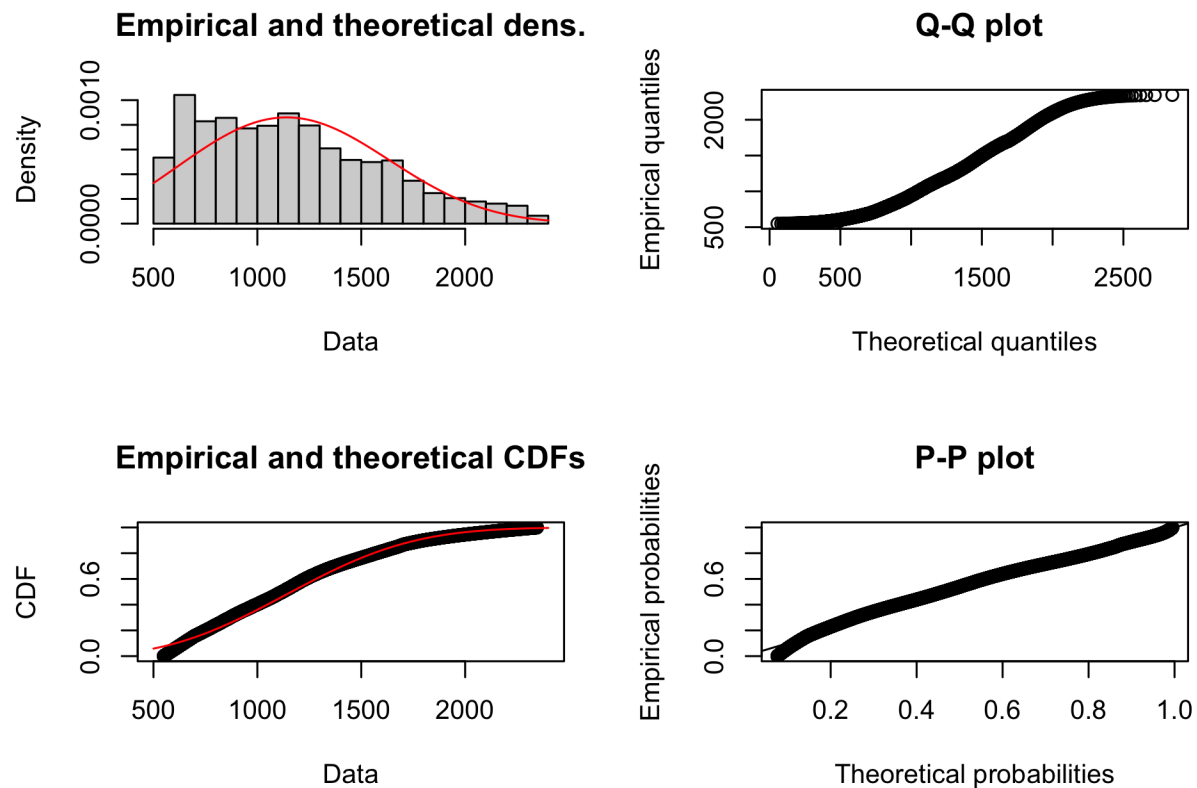
```
y_ch2 <- dtrunc(x, spec = "gamma", a = 400, b = 3000, shape = 7.4, rate = 0.006)
```

```
x_ch2_norm <- data.frame(cbind(x,y_ch2))

x_ch2_norm %>%
  ggplot(aes(x=x,y=y_ch2)) +
  geom_line(color= "red") +
  ggtitle("Modeled distribution (Switzerland / School 2)")+#plot density
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  theme_bw() +
  scale_y_continuous(expand = expansion(mult = c(0, 0.05)), limits = c(0, 0.004)) +
  scale_x_continuous(limits = c(400, 3000))
```



```
ch2_fit_weibull <- fitdist(x_ch2, "weibull") #different fitting function
plot(ch2_fit_weibull) #plots comparison
```



```
co2_distr_ch2 <- data.frame(co2 = seq(400, 3000, .1)) %>%
  mutate(prob = dtrunc(x, spec = "norm", a = 400, b = 3000, mean = 933.5, sd = 389.16))

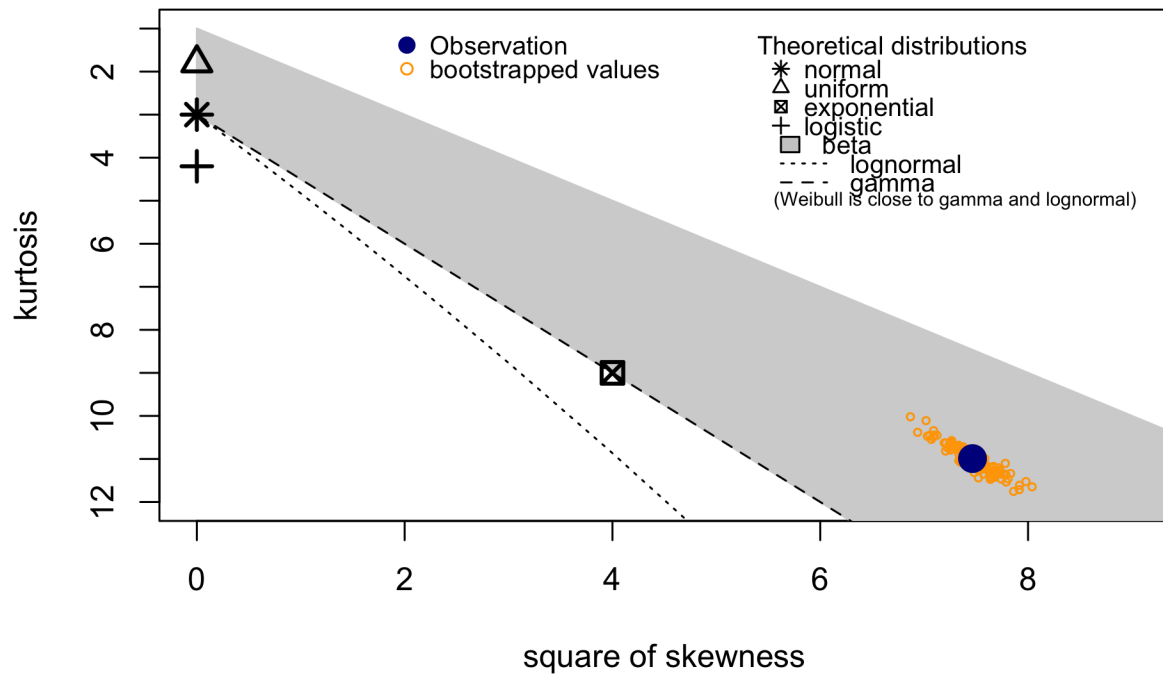
sample_co2_ch2 <- sample(co2_distr_ch2$co2, 10000, replace = TRUE, prob = co2_distr_ch2$prob) #sample
sample_f_ch2 <- tibble(co2 = sample_co2_ch2, f = ((co2-C_o)/C_a)/1000000) %>%
  dplyr::select(-co2)

#tanzania
x_tz <- loessData_tz %>%
  pull(y)

x_tz <- as.numeric(x_tz)

descdist(x_tz, discrete = FALSE, boot = 100) #gamma fits ok
```

Cullen and Frey graph



```
## summary statistics
## -----
## min: 460.1963 max: 1497.719
## median: 600.4938
## mean: 644.2189
## estimated sd: 173.3067
## estimated skewness: 2.732294
## estimated kurtosis: 10.99144
```

```
fitdistr(x_tz, "t") #get parameters
```

```
##           m           s           df
## 593.94973719 54.89374967 1.55018887
## ( 1.02943955) ( 1.01108907) ( 0.04266288)
```

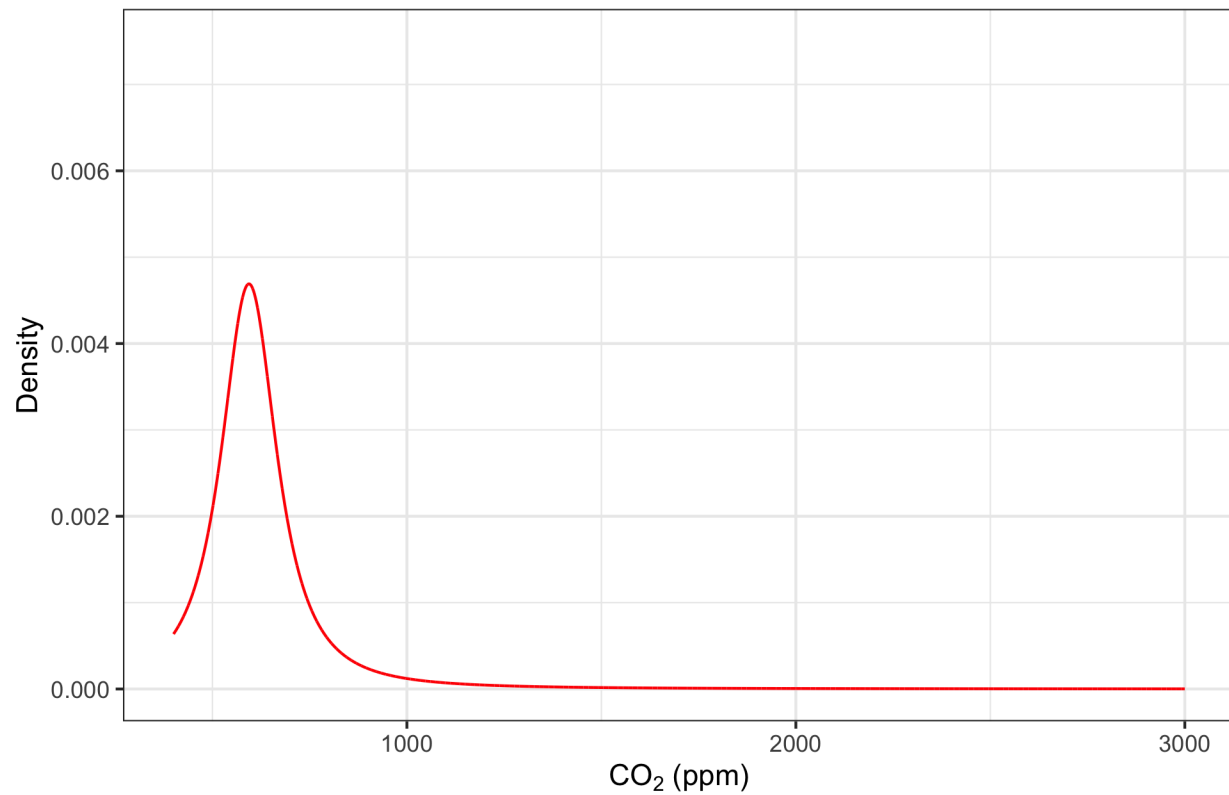
```
y_tz <- dtrunc(x, spec = "st", a = 400, b = 3000, mu = 593.95, sigma = 80, nu = 1.55)
```

```
x_tz_gamma <- data.frame(cbind(x,y_tz))
```

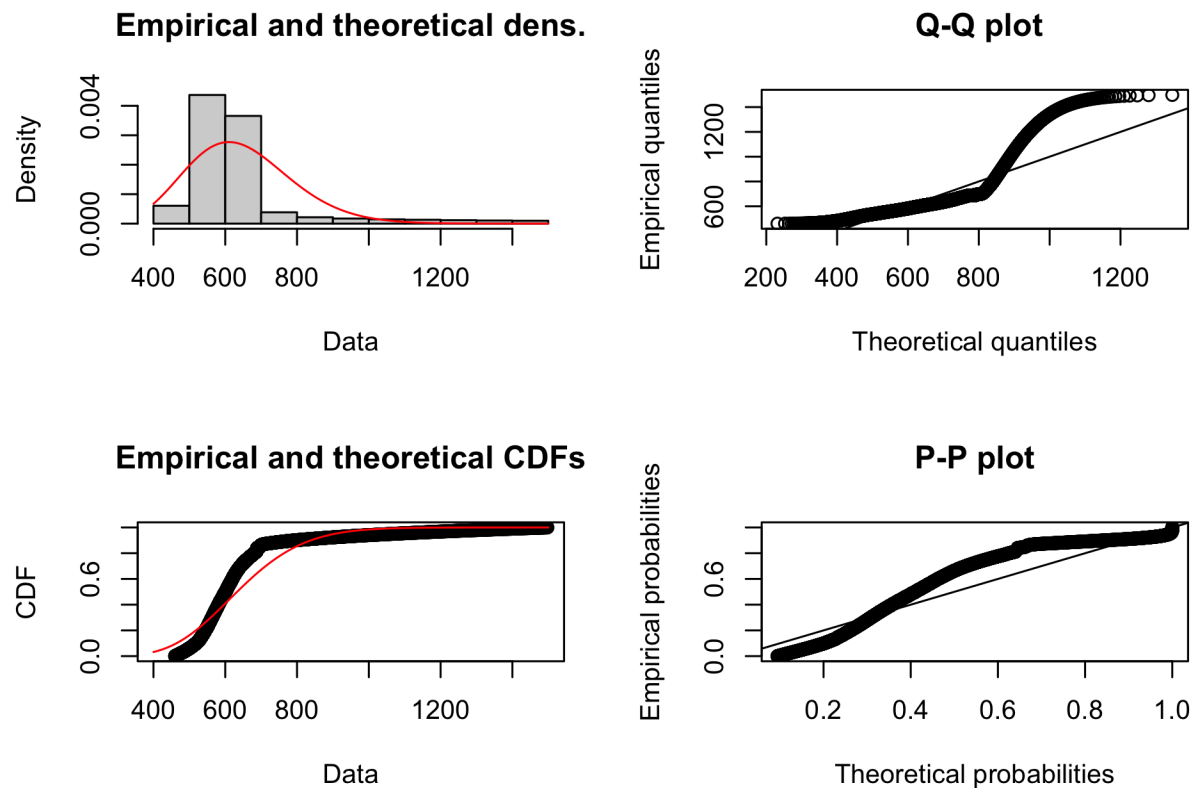
```
x_tz_gamma %>%
  ggplot(aes(x=x,y=y_tz)) +
  geom_line(color= "red") +
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  ggtitle("Modeled distribution (Tanzania)") +
```

```
theme_bw() +  
scale_y_continuous(limits = c(0, 0.0075)) +  
scale_x_continuous(limits = c(400, 3000))
```

Modeled distribution (Tanzania)



```
tz_fit_gamma <- fitdist(x_tz, "gamma") #different fitting function  
plot(tz_fit_gamma) #plots comparison
```



```
co2_distr_tz <- data.frame(co2 = seq(400, 3000, .1)) %>%
  mutate(prob = dtrunc(x, spec = "gamma", a = 400, b = 3000, shape = 18.2, rate = 0.03))

sample_co2_tz <- sample(co2_distr_tz$co2, 10000, replace = TRUE, prob = co2_distr_tz$prob) #sample
sample_f_tz <- tibble(co2 = sample_co2_tz, f = ((co2-C_o)/C_a)/1000000) %>%
  dplyr::select(-co2)
```

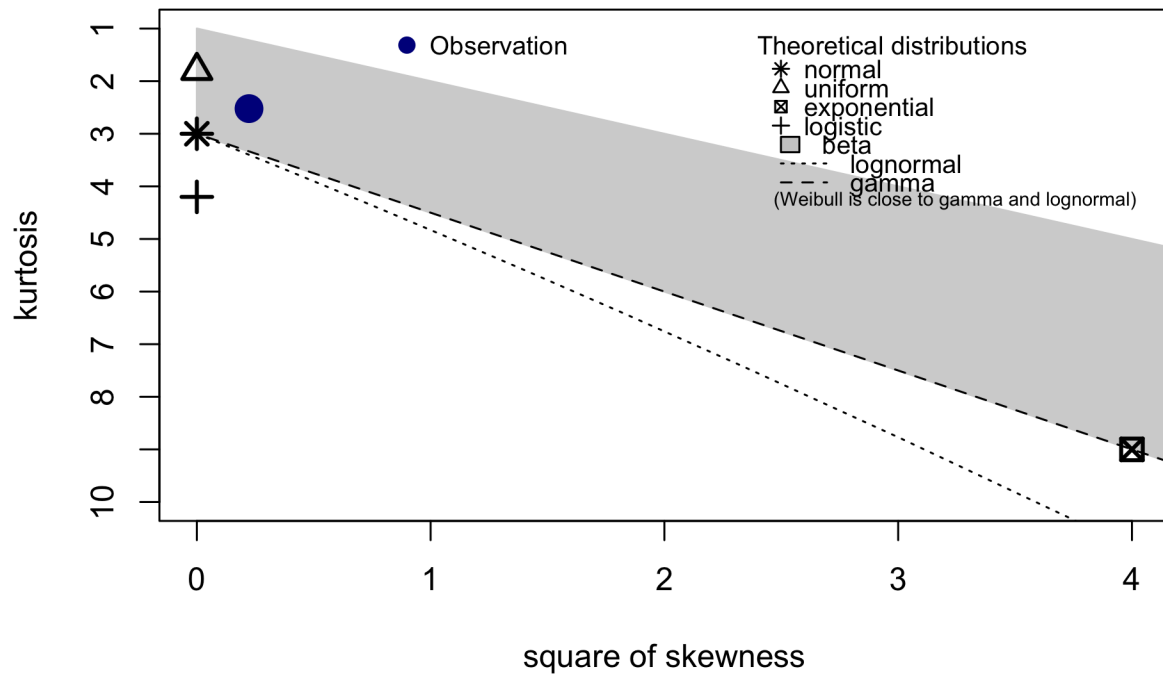
```
#south africa

x_sa <- loessData_sa %>%
  pull(y)

x_sa <- as.numeric(x_sa)

descdist(x_sa, discrete = FALSE) #normal/gamma fits ok -> after comparison --> gamma is better
```

Cullen and Frey graph



```
## summary statistics
## -----
## min: 418.1935   max: 2952.034
## median: 1375.813
## mean: 1457.681
## estimated sd: 596.2569
## estimated skewness: 0.4724284
## estimated kurtosis: 2.521861
```

```
fitdistr(x_sa, "gamma") #get parameters
```

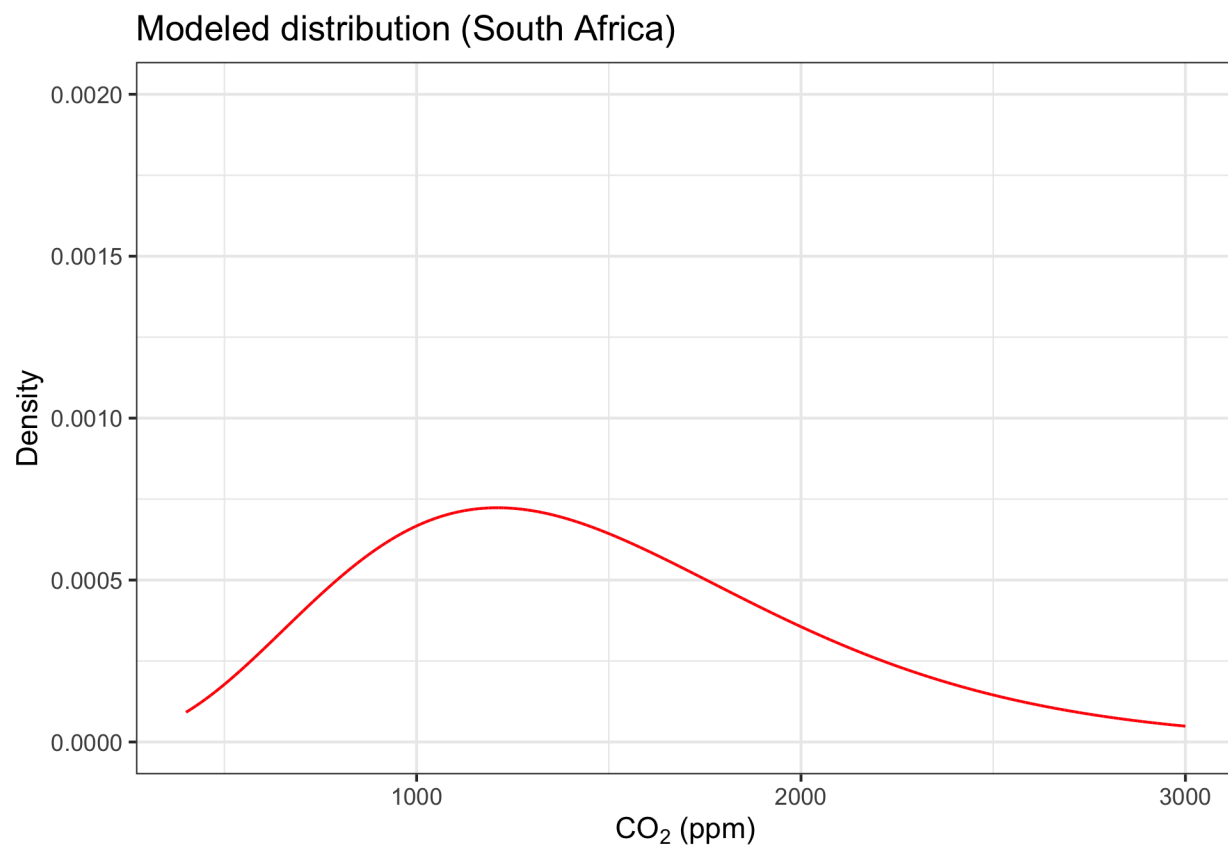
```
##           shape           rate
## 5.71414583537 0.00392003036
## (0.04226576130) (0.00002930492)
```

```
y_sa <- dtrunc(x, spec = "gamma", a = 400, b = 3000, shape = 5.72, rate = 0.0039)
```

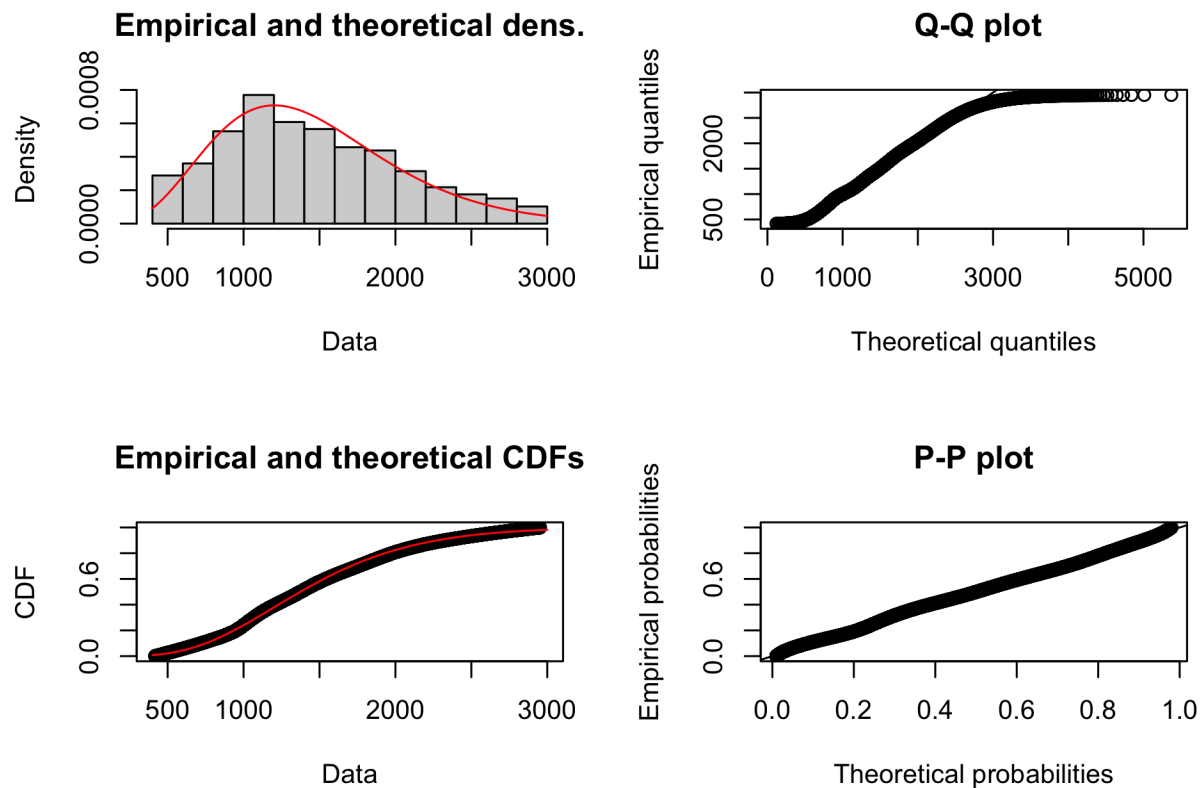
```
x_sa_gamma <- data.frame(cbind(x, y_sa))
```

```
x_sa_gamma %>%
  ggplot(aes(x=x, y=y_sa)) +
  geom_line(color = "red") +
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  ggtitle("Modeled distribution (South Africa)") +
```

```
theme_bw() +  
scale_y_continuous(limits = c(0, 0.002)) +  
scale_x_continuous(limits = c(400, 3000))
```



```
sa_fit_gamma<- fitdist(x_sa, "gamma") #different fitting function  
plot(sa_fit_gamma) #plots comparison
```

```
co2_distr_sa <- data.frame(co2 = seq(400, 3000, .1)) %>%
  mutate(prob = dtrunc(x, spec = "gamma", a = 400, b = 3000, shape = 5.7, rate = 0.0039))

sample_co2_sa <- sample(co2_distr_sa$co2, 10000, replace = TRUE, prob = co2_distr_sa$prob) #sample
sample_f_sa <- tibble(co2 = sample_co2_sa, f = ((co2-C_o)/C_a)/1000000) %>%
  dplyr::select(-co2)
```

Quanta

I'll use the following studies for calculating the mean parameter:

Riley (1962): 130 patients, q: 1.25 Escombe (2008): 117 patients, q: 8.2 Nardell (1991) : 1 patients, q: 12.5
 Andrews (2014) : 571 patients, q: 0.89

```
q <- (1.25*130+8.2*117+12.5+0.89*571)/(130+117+1+571) #weighted mean from different studies

#Escombe Table 2
mean_one_inf <- mean(c(12,3,5.5,1.8,18,12)) #mean quanta of pers. which infected one pig
mean_two_inf <- mean(c(2.9,40)) #mean quanta of pers. which infected two pigs

q_inf_persons <- c(12,3,2.9,5.5,1.8,18,40,12,226,52,mean_two_inf,rep(mean_one_inf,11))
q_sample_total_norm <- c(q_inf_persons, rtruncnorm(117-length(q_inf_persons), a = 0, mean = 1, sd = 1))

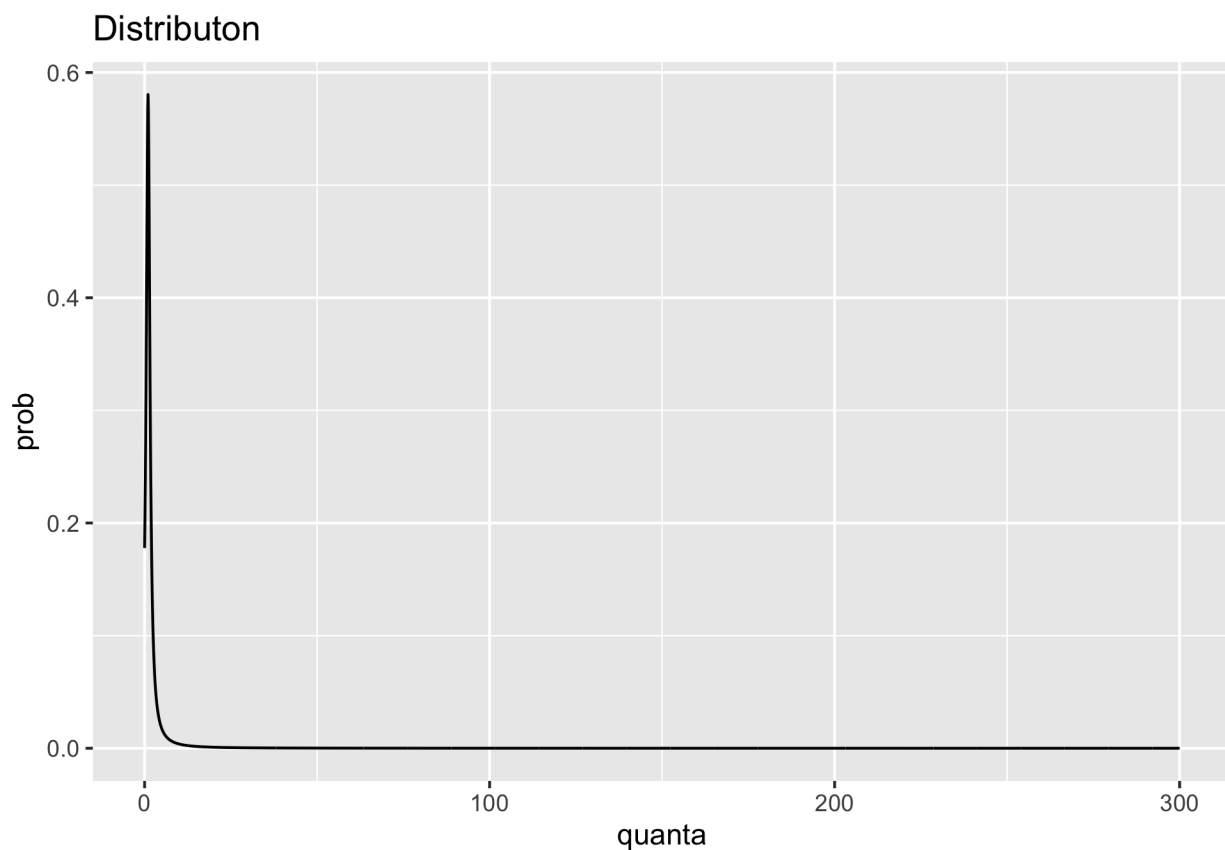
fitdistr(q_sample_total_norm, "t") #get parameters
```

```
##           m           s           df
##  1.10137081  0.65998949  0.89845036
##  (0.10037885) (0.09137468) (0.12845296)

dq <- function(x) {
  dtrunc(x, spec = "st", a = 0, b = 300, mu = 1, sigma = 0.67, nu = 0.90) #parameters from fitdistr
}

rq_distr <- data.frame(quanta = seq(0, 300, .1)) %>%
  mutate(prob = dq(quanta))

rq_distr %>%
  ggplot(aes(x = quanta, y = prob)) +
  geom_line() +
  ggtitle("Distributon") #Distribution of the quanta model
```



```
sample_q <- sample(rq_distr$quanta, 10000, replace = TRUE, prob = rq_distr$prob) #sample q for calculat

tb_sample <- tibble(sample = sample_q) %>%
  count(sample > 10) %>%
  mutate(prob = n/10000) #what is the ratio of datapoints over 10?

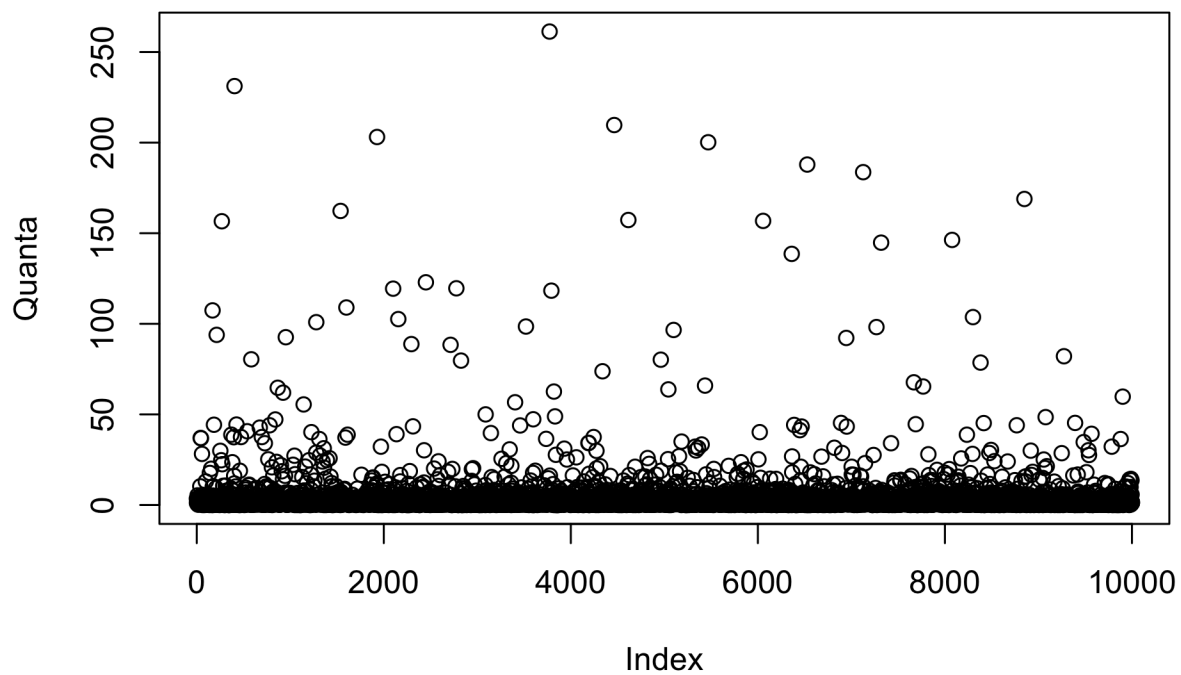
q_under_10 <- tb_sample[1,3, drop = TRUE]
q_under_10 # approx. 96% of data points are below 10
```

```
## [1] 0.9645
```

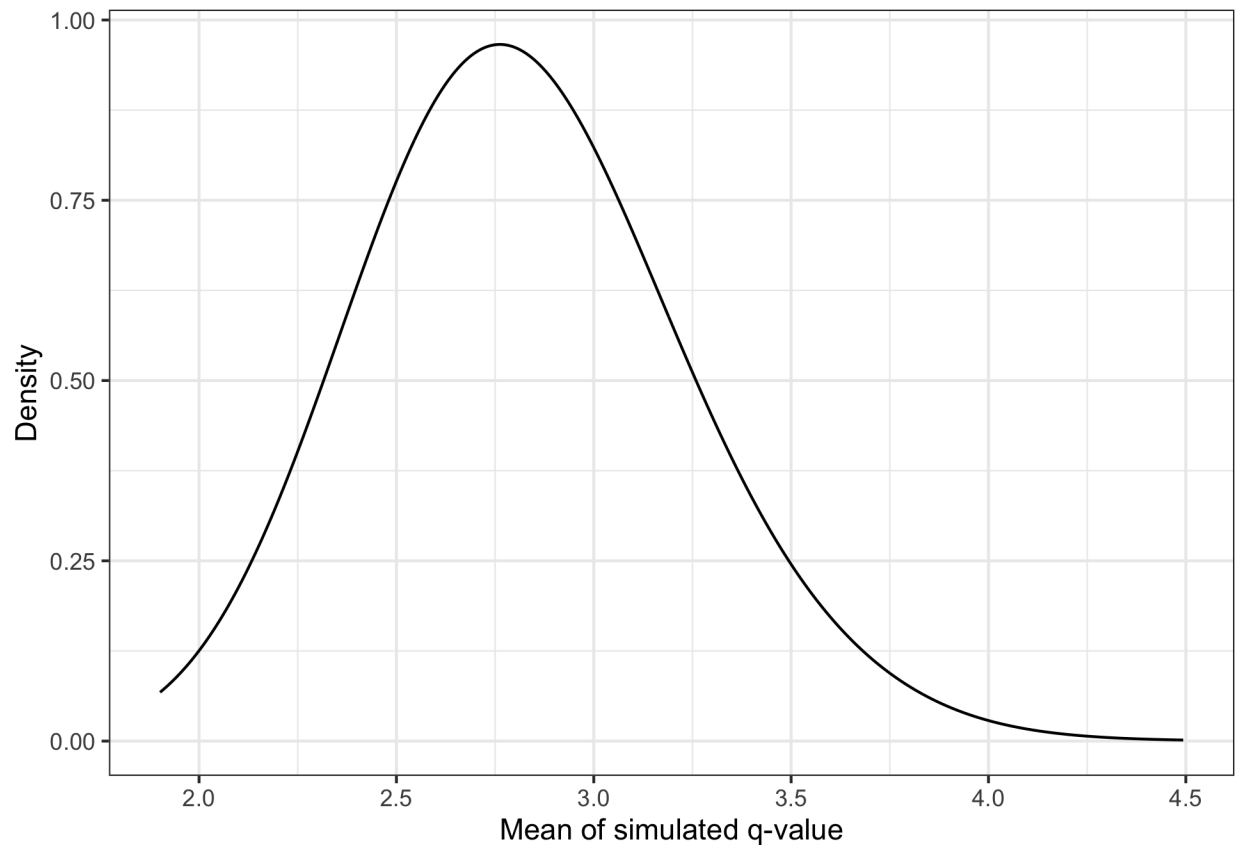
```
mean(sample_q) #mean of the sample should be between 2 and 3
```

```
## [1] 2.71343
```

```
plot(sample_q, ylab="Quanta")
```



```
thousand_sample <- replicate(10000, mean(sample(rq_distr$quanta, 1000, replace = TRUE, prob = rq_distr$  
#for calculating the distribution of the mean over lots of simulations  
ggplot() + geom_density(mapping = aes(thousand_sample), alpha = .2, kernel = "gaussian", adjust = 5) +  
  xlab("Mean of simulated q-value") +  
  ylab("Density") +  
  theme_bw()
```



#Mean is in the desired area

Rest of the parameters

```
#n
n_ch <- 20
n_sa <- 30 #Powerpoint
n_tz <- 50 #Powerpoint

#I
prev_ch <- 4.14/100000
#https://www.bag.admin.ch/bag/de/home/zahlen-und-statistiken/zahlen-zu-infektionskrankheiten.exturl.htm
prev_ch_y <- 8.23/100000

age_group <- sum(c(2566719, 2534956, 2351752, 2327273))
#https://www.statista.com/statistics/1330839/population-of-south-africa-by-age-group-and-gender/
prev_sa <- 513/100000
prev_sa_y <- (20000+12000)/age_group # tendenziell zu hoch da 15-25
#https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_entity_type=%22country%22&lan=%22EN%22&iso2
# Ansteckungen in Altersgruppe 5-24 durch Population in dieser Altersgruppe (nicht 15-24 genommen, da s
prev_tz <- 208/100000
#https://data.worldbank.org/indicator/SH.TBS.INCD?locations=TZ
prev_tz_y <- 208/100000
```

#agegroup_tz #https://www.nbs.go.tz/nbs/takwimu/census2012/atlas/POPULATION_AGED_15-24.pdf

```
I_ch <- prev_ch*n_ch #prevalence per class (per year)
I_ch_y <- prev_ch_y*n_ch
I_sa <- prev_sa*n_sa
I_sa_y <-prev_sa_y*n_sa
I_tz <- prev_tz*n_tz
I_tz_y <- prev_tz_y*n_tz
```

```
month <- 8*5*4
year <- 8*5*4*10
```

#preparing datasets for plotting

```
df_ch1 <- tibble(school = c(rep("school 1", 10000)), f = sample_f_ch1, q = sample_q) %>%
  mutate(P_year = 1 - exp(-(f*I_ch*q*year)/n_ch)) %>%
  mutate(P_month = 1 - exp(-(f*I_ch*q*month)/n_ch)) %>%
  mutate(P_year_y = 1- exp(-(f*I_ch_y*q*year)/n_ch)) %>%
  mutate(P_year_one = 1 - exp(-(f*1*q*year)/n_ch)) %>%
  mutate(P_month_one = 1 - exp(-(f*1*q*month)/n_ch)) %>%
  mutate(five_quantil = unname(quantile(P_year$f, probs = seq(0, 1, 1/20)))[2]) %>%
  mutate(ninefive_quantil = unname(quantile(P_year$f, probs = seq(0, 1, 1/20)))[20])
```

```
df_ch2 <- tibble(school = c(rep("school 2", 10000)), f = sample_f_ch2, q = sample_q) %>%
  mutate(P_year = 1 - exp(-(f*I_ch*q*year)/n_ch)) %>%
  mutate(P_month = 1 - exp(-(f*I_ch*q*month)/n_ch)) %>%
  mutate(P_year_y = 1- exp(-(f*I_ch_y*q*year)/n_ch)) %>%
  mutate(P_year_one = 1 - exp(-(f*1*q*year)/n_ch)) %>%
  mutate(five_quantil = unname(quantile(P_year$f, probs = seq(0, 1, 1/20)))[2]) %>%
  mutate(ninefive_quantil = unname(quantile(P_year$f, probs = seq(0, 1, 1/20)))[20])
```

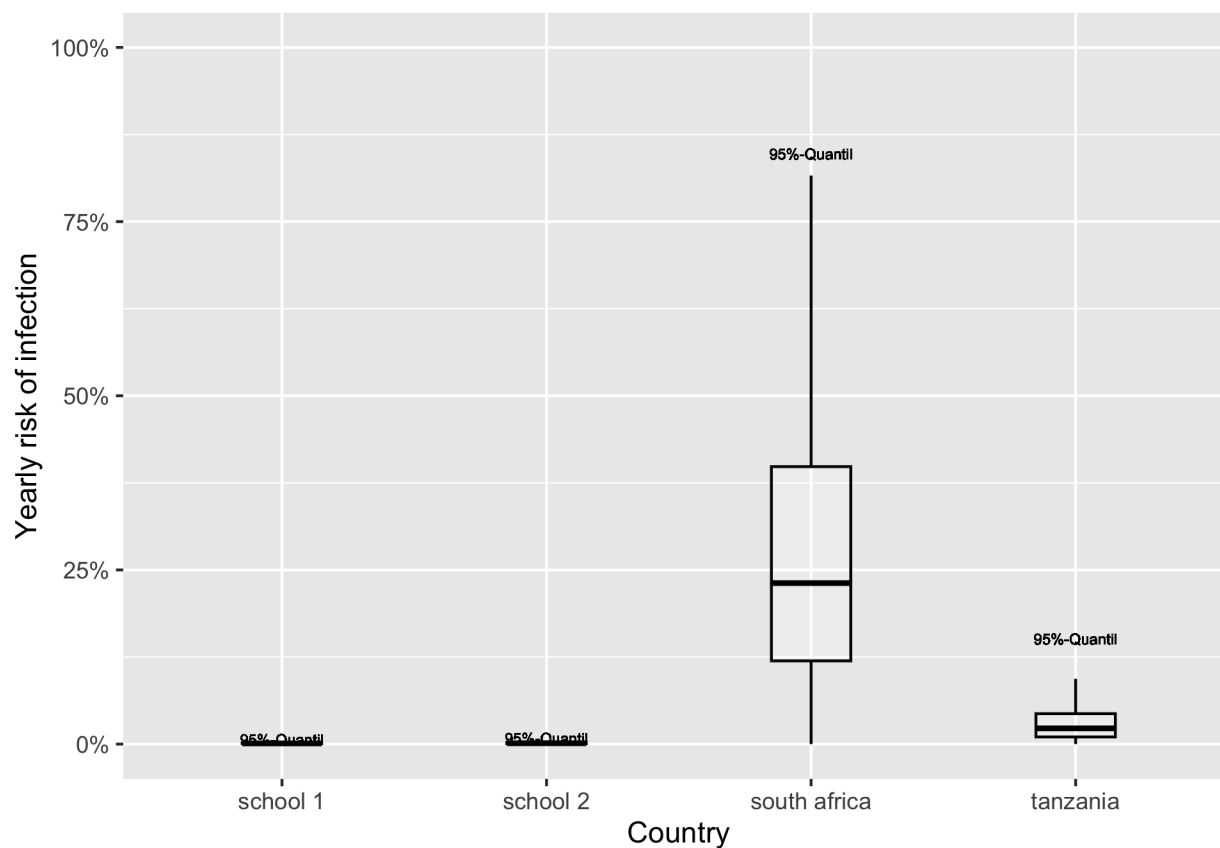
```
df_tz <- tibble(school = c(rep("tanzania", 10000)), f = sample_f_tz, q = sample_q) %>%
  mutate(P_year = 1 - exp(-(f*I_tz*q*year)/n_tz)) %>%
  mutate(P_month = 1 - exp(-(f*I_tz*q*month)/n_tz)) %>%
  mutate(P_year_y = 1- exp(-(f*I_tz_y*q*year)/n_tz)) %>%
  mutate(P_year_one = 1 - exp(-(f*1*q*year)/n_tz)) %>%
  mutate(five_quantil = unname(quantile(P_year$f, probs = seq(0, 1, 1/20)))[2]) %>%
  mutate(ninefive_quantil = unname(quantile(P_year$f, probs = seq(0, 1, 1/20)))[20])
```

```
df_sa <- tibble(school = c(rep("south africa", 10000)), f = sample_f_sa, q = sample_q) %>%
  mutate(P_year = 1 - exp(-(f*I_sa*q*year)/n_sa)) %>%
  mutate(P_month = 1 - exp(-(f*I_sa*q*month)/n_sa)) %>%
  mutate(P_year_y = 1- exp(-(f*I_sa_y*q*year)/n_sa)) %>%
  mutate(P_year_one = 1 - exp(-(f*1*q*year)/n_sa)) %>%
  mutate(five_quantil = unname(quantile(P_year$f, probs = seq(0, 1, 1/20)))[2]) %>%
  mutate(ninefive_quantil = unname(quantile(P_year$f, probs = seq(0, 1, 1/20)))[20])
```

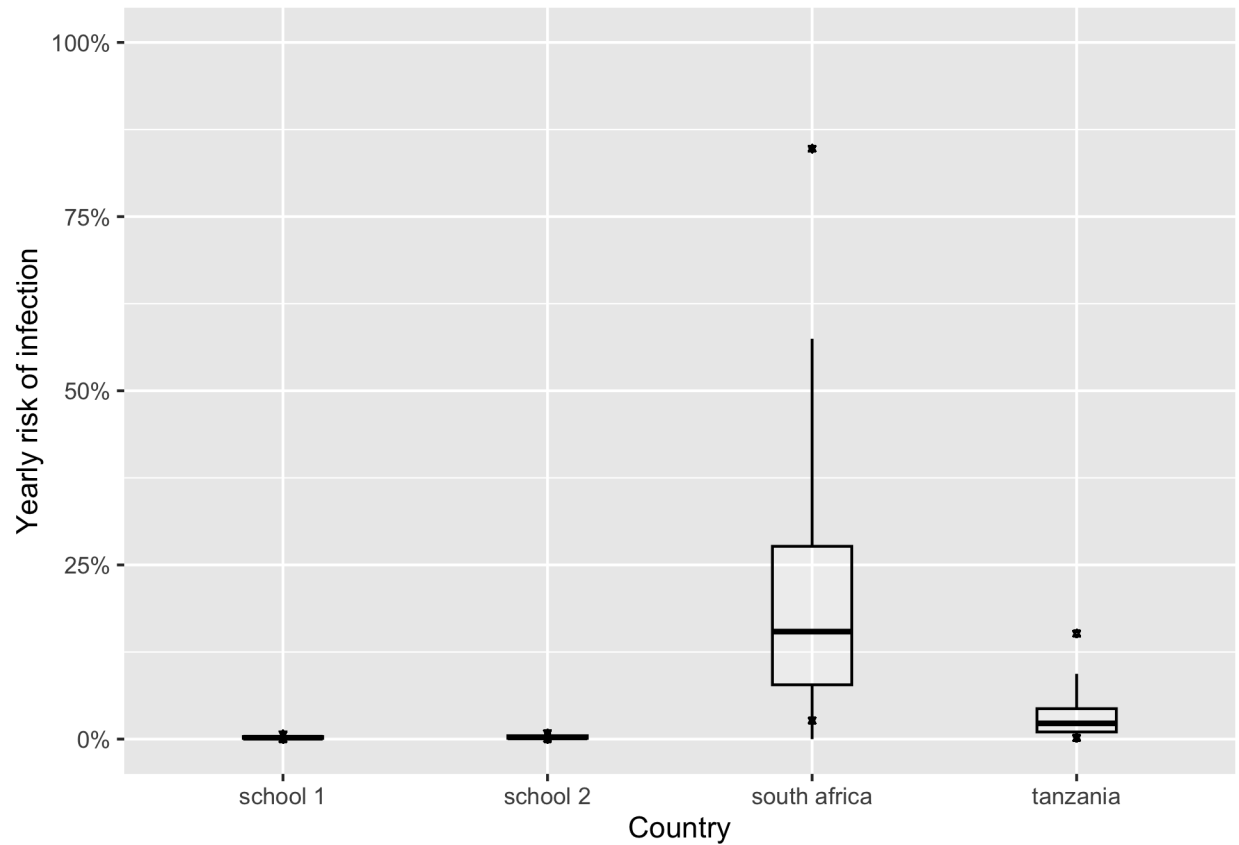
```
df_complet <- bind_rows(df_ch1, df_ch2, df_sa, df_tz)
```

Plots of the transmission risk

```
df_complet %>%
  ggplot(aes(x = school, y=P_year$f))+
  geom_boxplot(width=0.3, color="black", alpha=0.2, outlier.shape = NA) +
  scale_y_continuous(labels = scales::percent_format(scale = 100)) +
  geom_text(aes(x= school, y= ninefive_quantil), label = "95%-Quantil", show.legend = FALSE, size = 2)+
  xlab("Country") +
  ylab("Yearly risk of infection")
```



```
## prevalence for youth
df_complet %>%
  ggplot(aes(x = school, y=P_year_y$f))+
  geom_boxplot(width=0.3, color="black", alpha=0.2, outlier.shape = NA) +
  scale_y_continuous(labels = scales::percent_format(scale = 100)) +
  geom_point(aes(x= school, y= five_quantil), show.legend = FALSE, size = 0.5, shape = 11)+
  geom_point(aes(x= school, y= ninefive_quantil), show.legend = FALSE, size = 0.5, shape = 11)+
  xlab("Country") +
  ylab("Yearly risk of infection")
```



Now I will compare the risks of infection, assuming that the prevalence is the same for every country and also assuming that the class size is the same. The prevalence per country is not used. This is to highlight the influence of air quality.

```
df_complet %>%
  ggplot(aes(x = school, y=P_year_one$f, colour = school)) +
  geom_boxplot(width=0.3, color="black", alpha=0.2) +
  scale_y_continuous(labels = scales::percent_format(scale = 100)) +
  xlab("Country") +
  ylab("Yearly risk of infection")
```

