

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)
```

Data

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

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

sa <- satz %>%
  filter(country == "South Africa") %>%
  filter(co2 < 3000) %>%
  filter(co2 > 400) #south africa data

tz <- satz %>%
  filter(country == "Tanzania") %>%
  filter(co2 < 3000) %>%
  filter(co2 > 400) #tanzania data
```

Methods

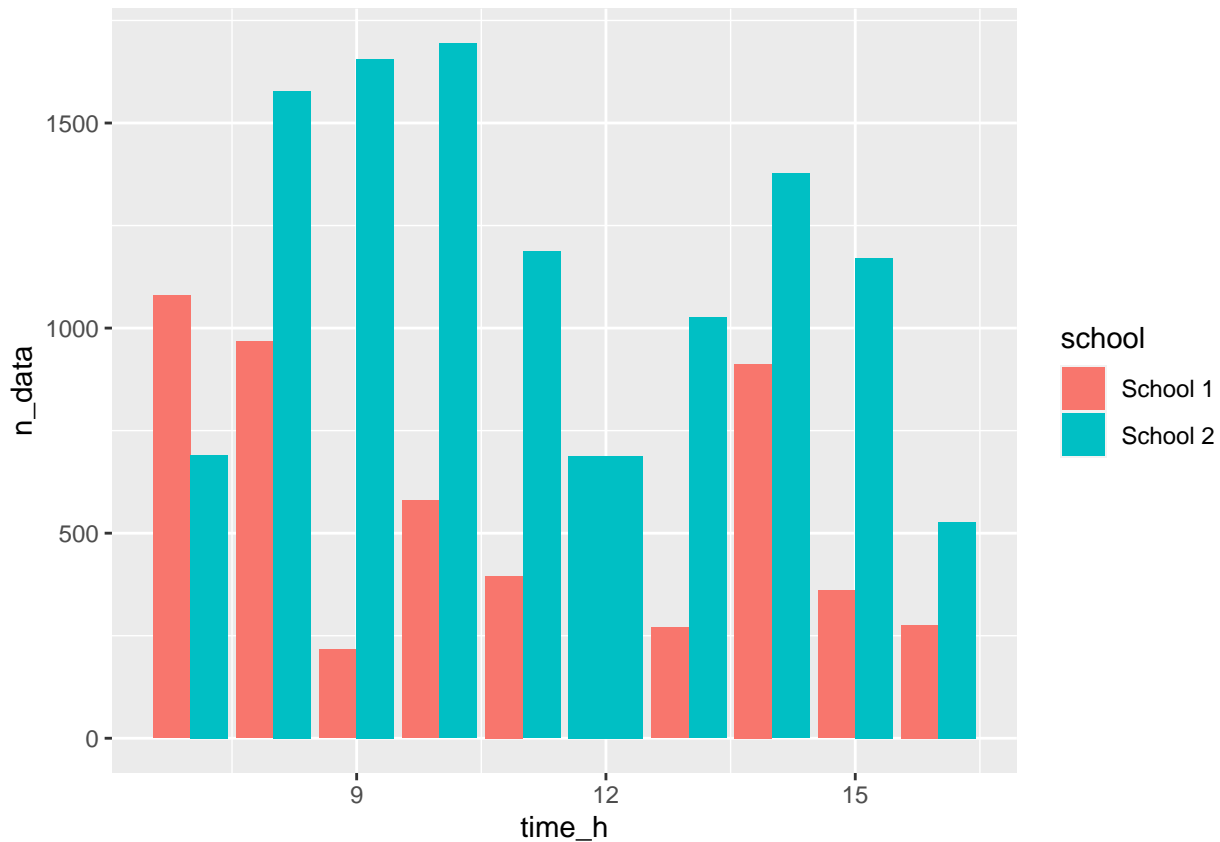
Indoor Co2 concentration * mean or distribution from data * C_o := Outdoor Co2 concentration * from literature https://www.fsis.usda.gov/sites/default/files/media_file/2020-08/Carbon-Dioxide.pdf * C_a := Vol-

ume 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

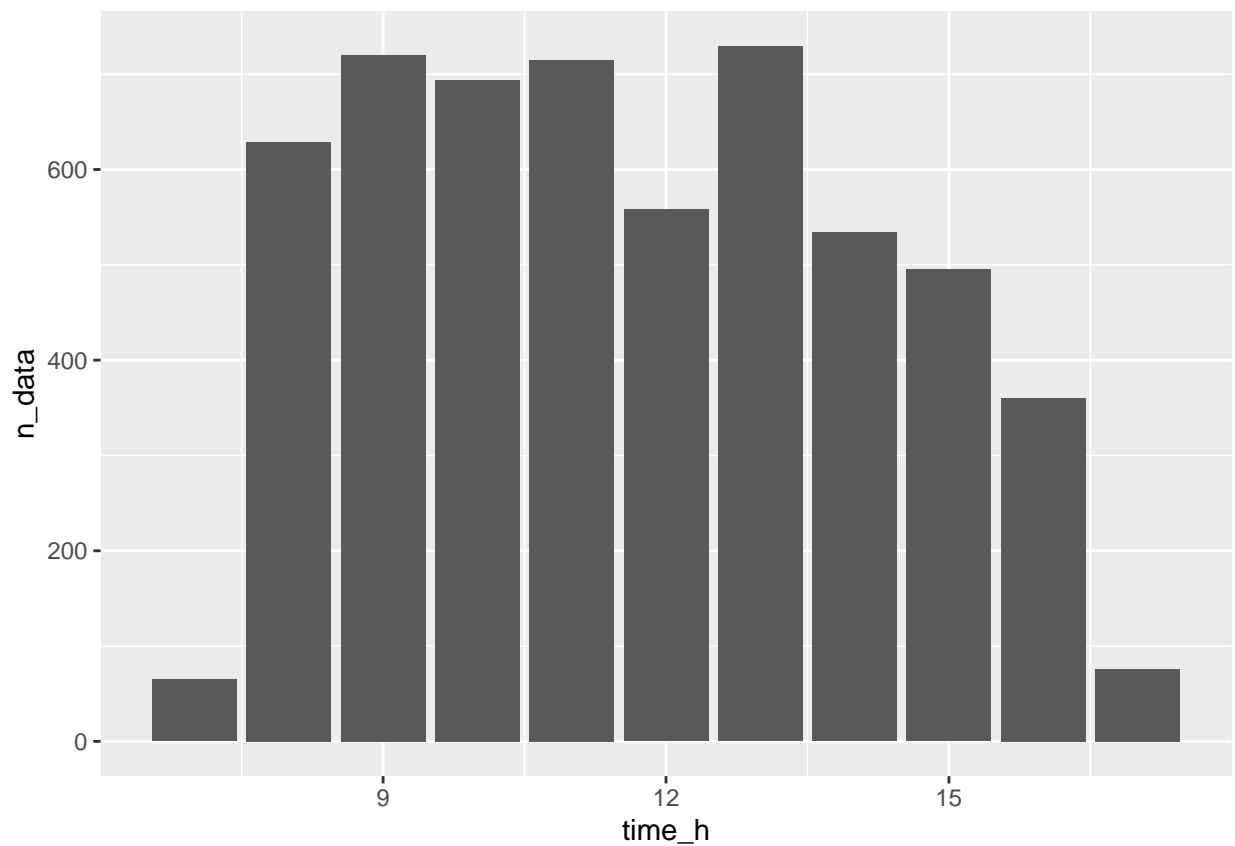
```
ch_hourly <- ch %>%
  mutate(time_h = hour(time)) %>%
  group_by(school, time_h) %>%
  summarize(mean = mean(co2),
             lower = quantile(co2, 0.25),
             upper = quantile(co2, 0.75),
             n_data = n()) %>%
  ungroup()

ch_hourly %>%
  ggplot(aes(x = time_h, y = n_data, fill = school)) +
  geom_bar(stat = "identity", position = position_dodge())
```



```
# 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())
```



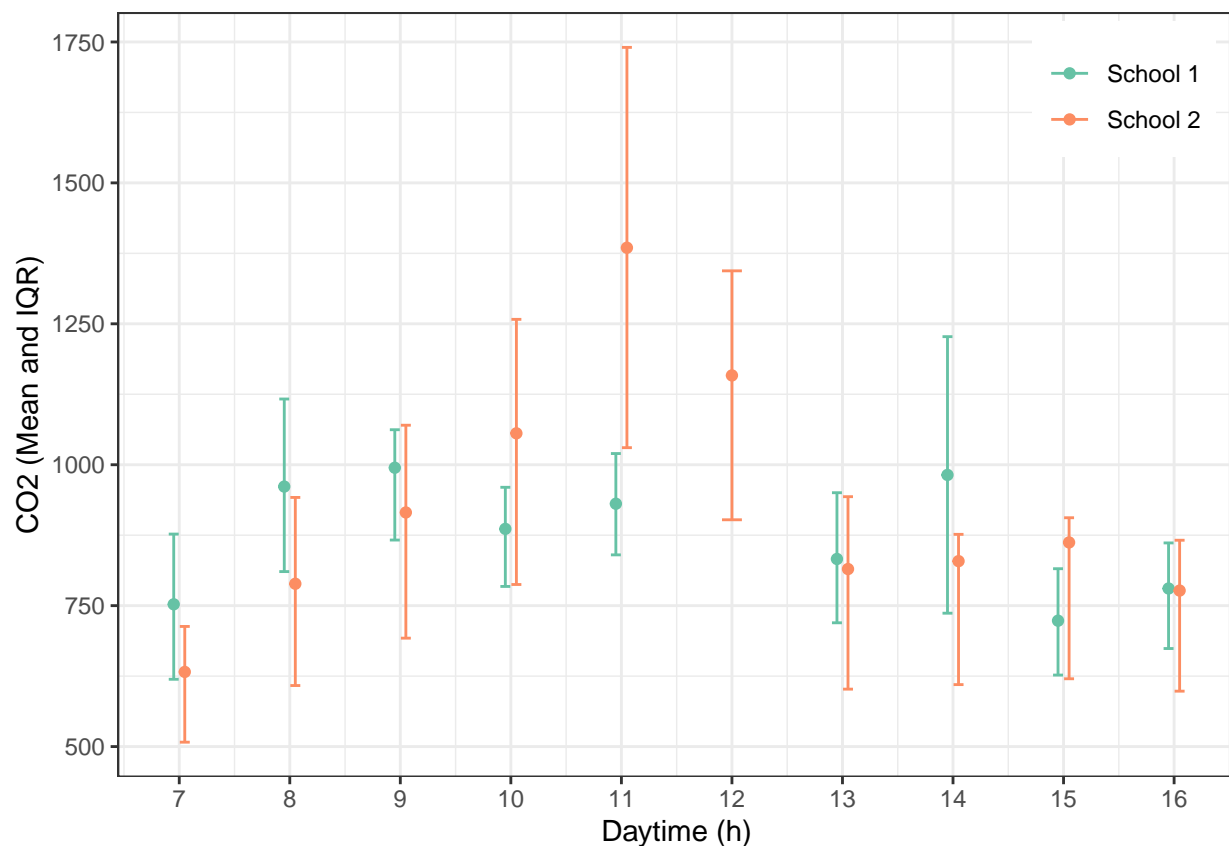
```
# data is measured throughout the day in south africa
```

```
# no time data for south africa
```

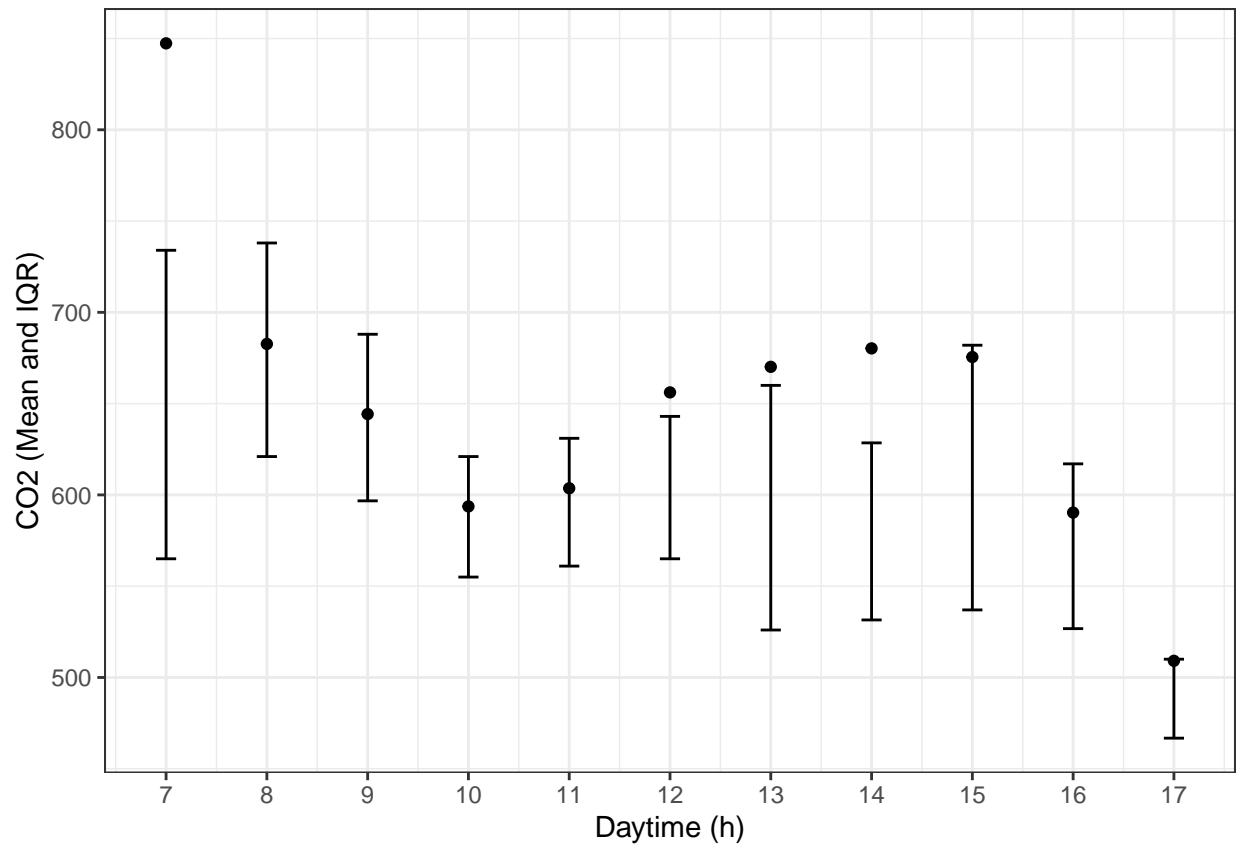
Analysis

Co2 over time

```
ch_hourly %>%
  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())
```

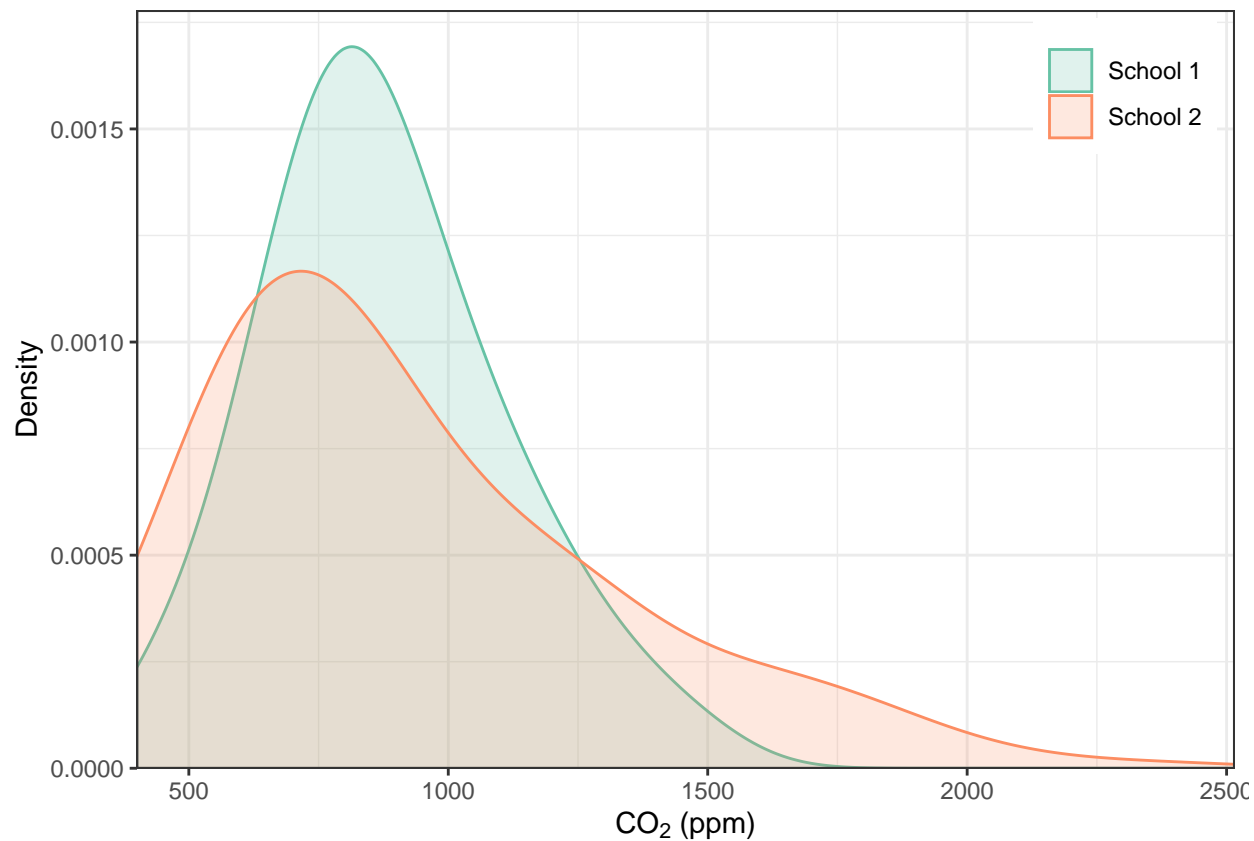


```
tz_hourly %>%
  ggplot(aes(x = time_h)) +
  geom_errorbar(aes(ymin = lower, ymax = upper), width = .2, position = position_dodge2(width = .2)) +
  geom_point(aes(y = mean), position = position_dodge2(width = .2)) +
  scale_color_brewer(palette = "Set2") +
  scale_x_continuous(breaks = seq(7, 17,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

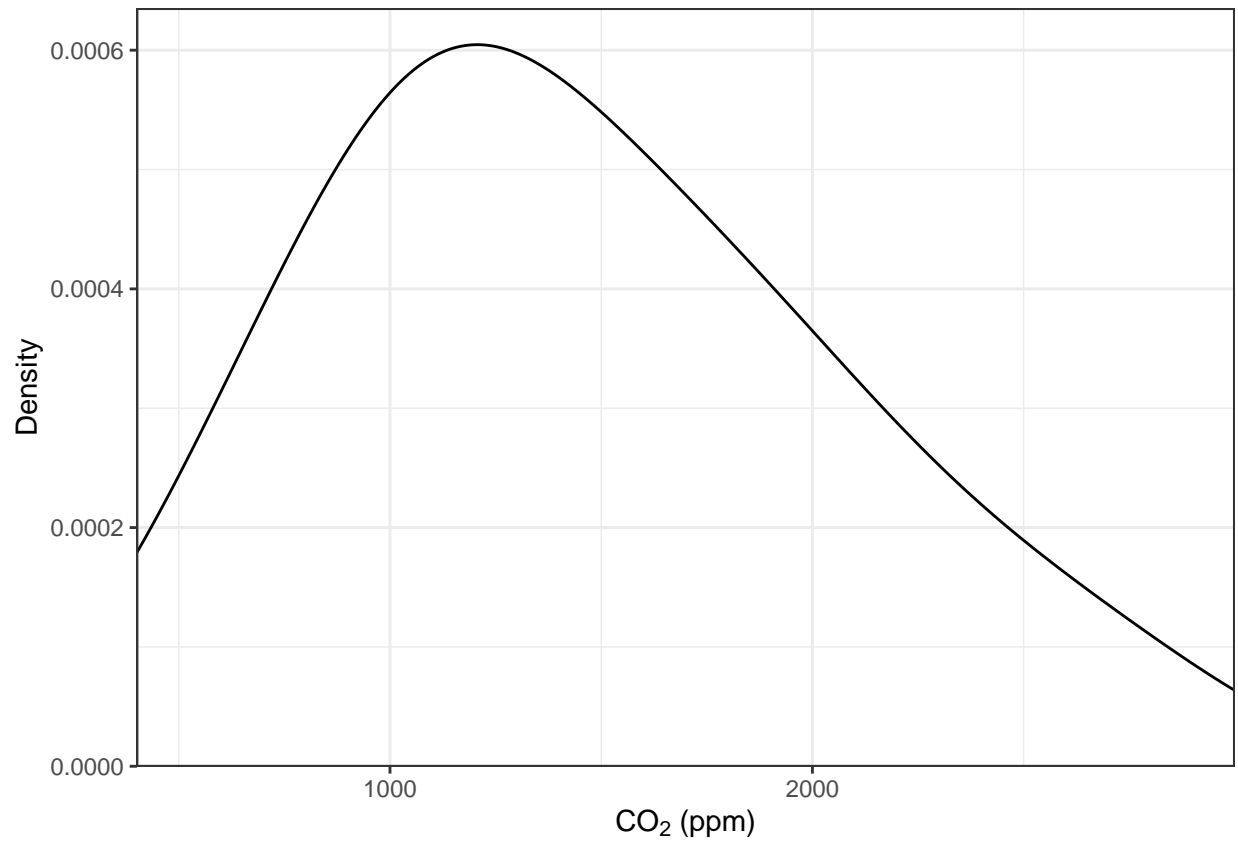


Co2 distribution

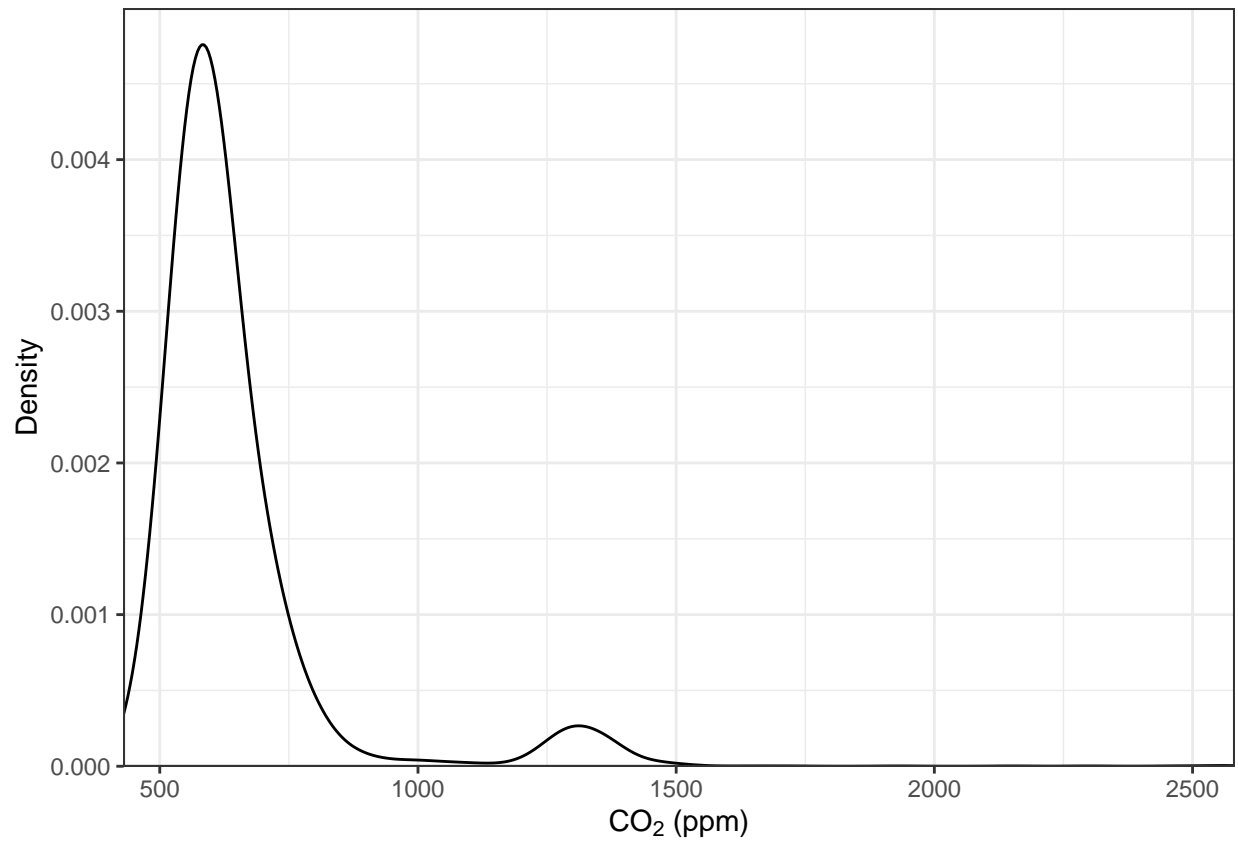
```
ch %>%
  ggplot(aes(x = co2, color = school, fill = school)) +
  geom_density(alpha = .2, kernel = "gaussian", adjust = 3) +
  scale_color_brewer(palette = "Set2") +
  scale_fill_brewer(palette = "Set2") +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.05))) +
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank())
```



```
sa %>%
  ggplot(aes(x = co2)) +
  geom_density(alpha = .2, kernel = "gaussian", adjust = 4) +
  scale_color_brewer(palette = "Set2") +
  scale_fill_brewer(palette = "Set2") +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.05))) +
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank())
```



```
tz %>%
  ggplot(aes(x = co2)) +
  geom_density(alpha = .2, kernel = "gaussian", adjust = 3) +
  scale_color_brewer(palette = "Set2") +
  scale_fill_brewer(palette = "Set2") +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.05))) +
  labs(x = expression(CO[2]*" (ppm)"), y = "Density") +
  theme_bw() +
  theme(legend.position = c(0.9,0.9), legend.title = element_blank())
```



Distribution co2

```
library(fitdistrplus)

#C_a
C_a <- 0.038 #Rudnik

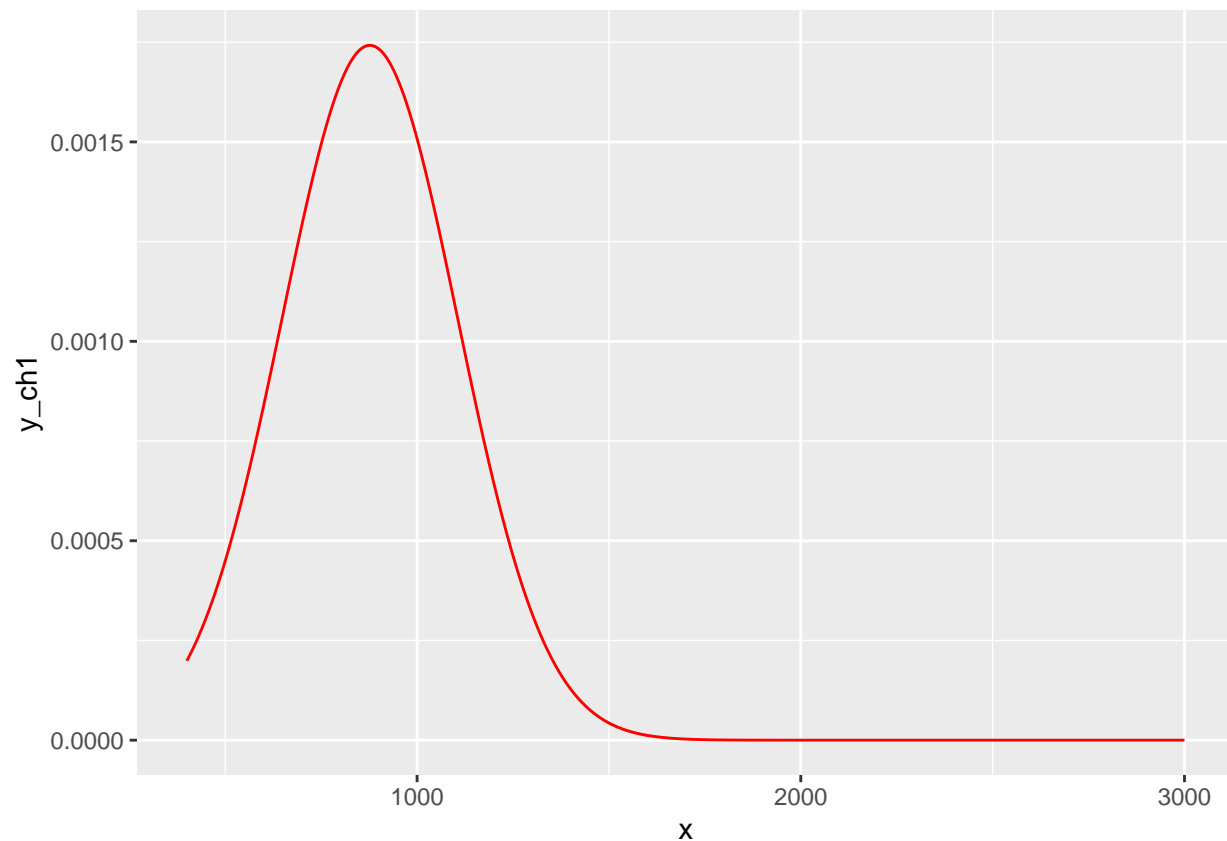
#C_o
C_o <- 400 #p.p.m (taking a higher estimate because higher values ar possible when a lot of traffic ect

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

#descdist(x_ch1, discrete = FALSE) #normal distribution fits well
#fitdistr(x_ch1, "normal") #get parameters

x <- seq(400, 3000, by = .1)
y_ch1 <- dnorm(x, mean = 877, sd = 229)
x_ch1_norm <- data.frame(cbind(x,y_ch1))

x_ch1_norm %>%
  ggplot(aes(x=x,y=y_ch1)) +
  geom_line(color= "red") #plot density
```

```
#ch1_fit_norm <- fitdist(x_ch1, "norm", lower=c(0,0)) #different fitting function
#plot(ch1_fit_norm) #plots comparison

co2_distr_ch1 <- data.frame(co2 = seq(400, 3000, .1)) %>%
  mutate(prob = dnorm(co2,877,299))

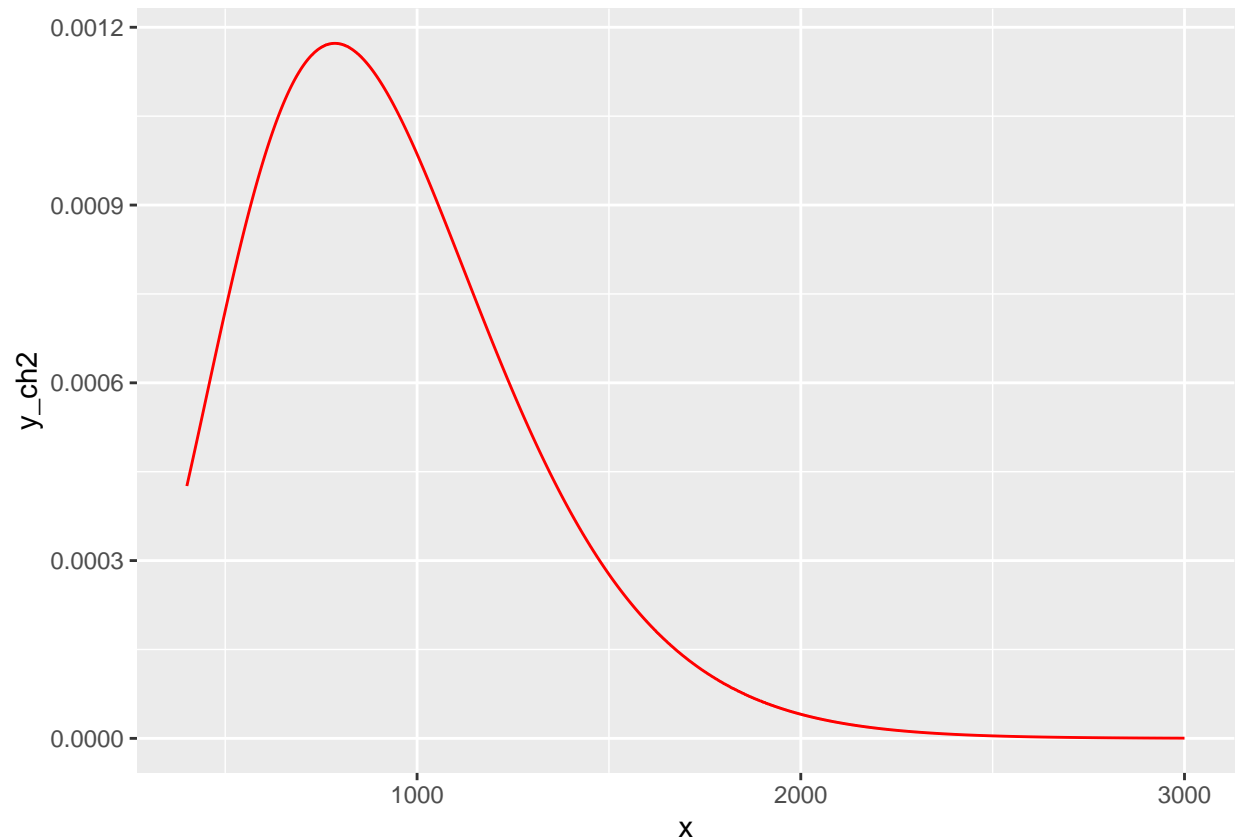
sample_co2_ch1 <- sample(co2_distr_ch1$co2, 1000, replace = TRUE, prob = co2_distr_ch1$prob) #sample
sample_f_ch1 <- tibble(co2 = sample_co2_ch1, f = ((co2-C_o)/C_a)/100000) %>%
  dplyr::select(-co2)

#school 2
x_ch2 <- ch %>%
  filter(school == "School 2") %>%
  pull(co2)

#descdist(x_ch2, discrete = FALSE) #gamma distribution fits well
#fitdistr(x_ch2, "gamma") #get parameters

y_ch2 <- dgamma(x, 6.5, 0.007)
x_ch2_gamma <- data.frame(cbind(x,y_ch2))

x_ch2_gamma %>%
  ggplot(aes(x=x,y=y_ch2)) +
  geom_line(color= "red") #plot density
```



```
#ch2_fit_gamma <- fitdist(x_ch2, "gamma") #different fitting function
#plot(ch2_fit_gamma) #plots comparison

co2_distr_ch2 <- data.frame(co2 = seq(400, 3000, .1)) %>%
  mutate(prob = dgamma(co2,6.5,0.007))

sample_co2_ch2 <- sample(co2_distr_ch2$co2, 1000, replace = TRUE, prob = co2_distr_ch2$prob) #sample
sample_f_ch2 <- tibble(co2 = sample_co2_ch2, f = ((co2-C_o)/C_a)/100000) %>%
  dplyr::select(-co2)

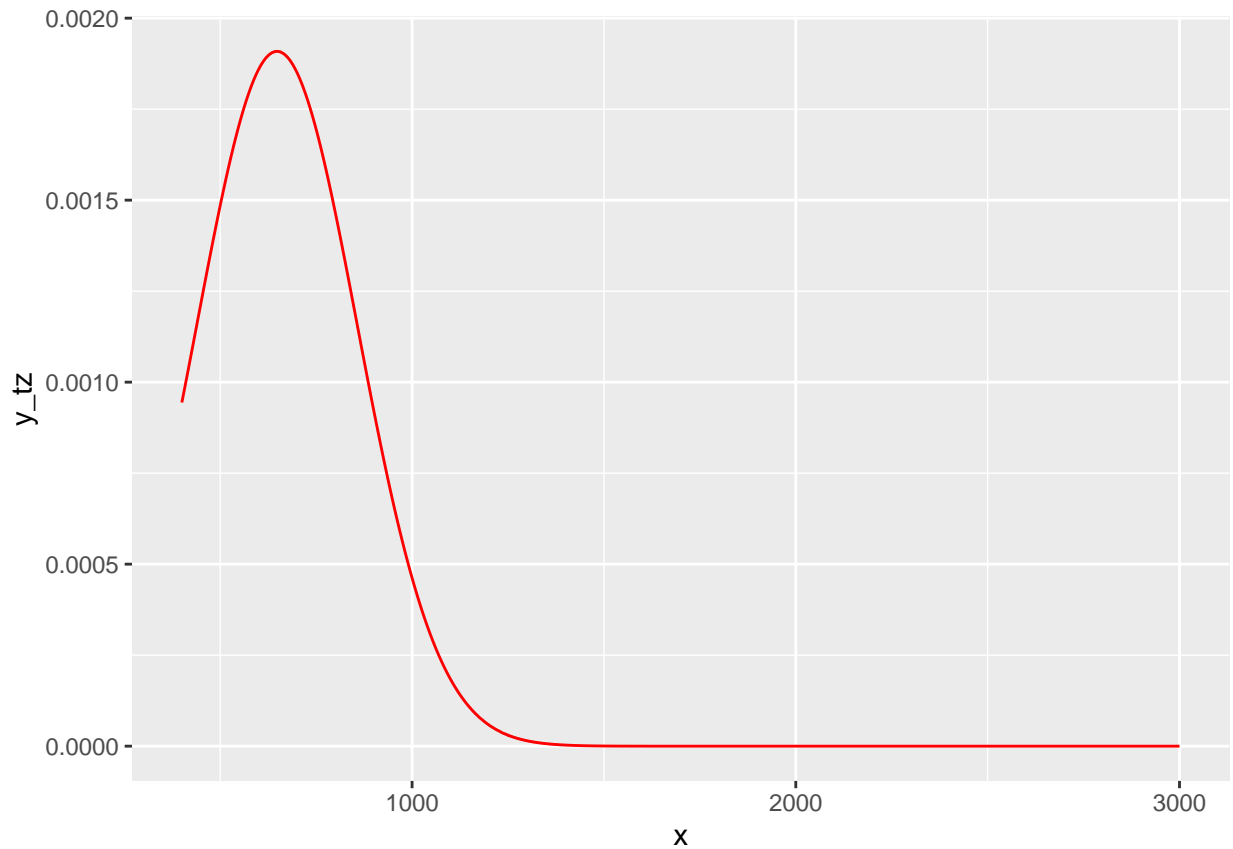
#tanzania

x_tz <- tz %>%
  pull(co2)

#descdist(x_tz, discrete = FALSE) #normal distribution fits well
#fitdistr(x_tz, "normal")

y_tz <- dnorm(x, mean = 648, sd = 209)
x_tz_norm <- data.frame(cbind(x,y_tz))

x_tz_norm %>%
  ggplot(aes(x=x,y=y_tz)) +
  geom_line(color= "red")
```



```

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

co2_distr_tz <- data.frame(co2 = seq(400, 3000, .1)) %>%
  mutate(prob = dnorm(co2,648,209))

sample_co2_tz <- sample(co2_distr_tz$co2, 1000, replace = TRUE, prob = co2_distr_tz$prob) #sample
sample_f_tz <- tibble(co2 = sample_co2_tz, f = ((co2-C_o)/C_a)/100000) %>%
  dplyr::select(-co2)

#south africa

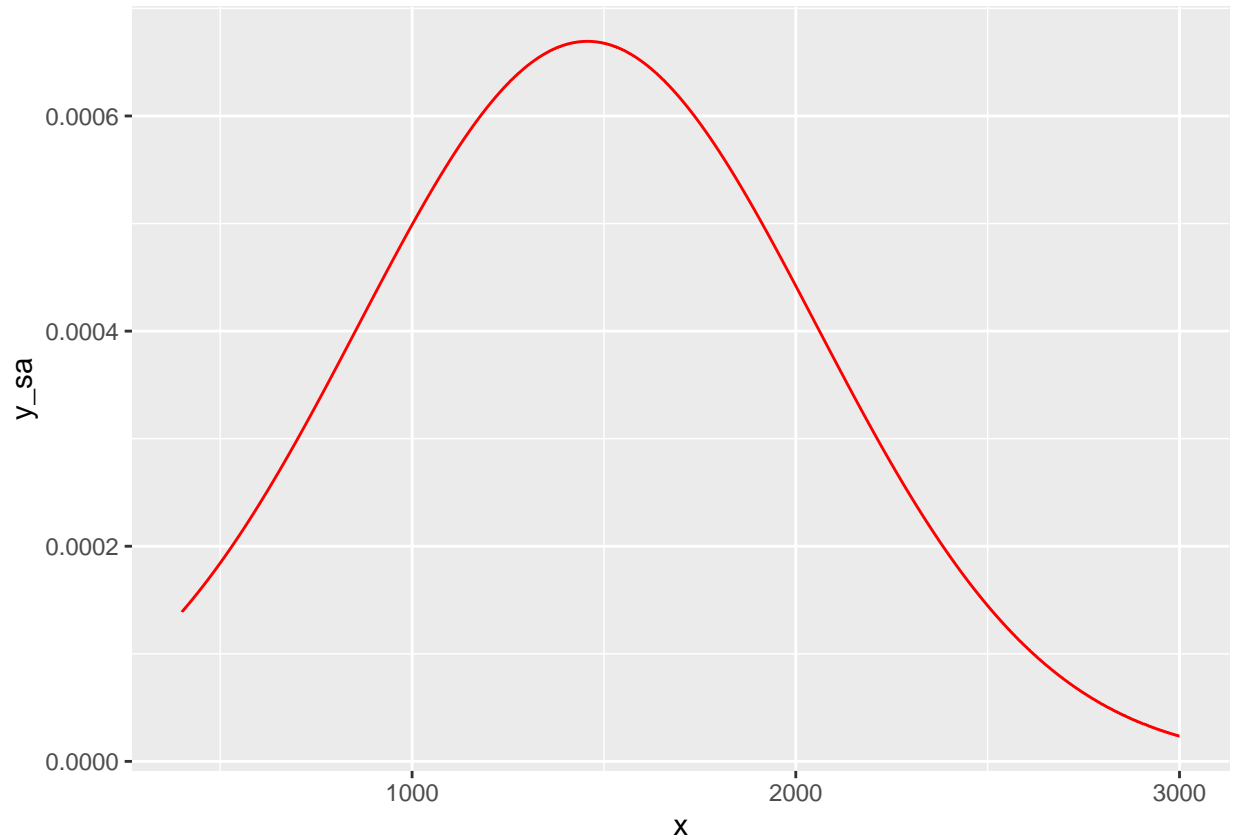
x_sa <- sa %>%
  pull(co2)

#descdist(x_sa, discrete = FALSE) #normal distribution fits well
#fitdistr(x_sa, "normal")

y_sa <- dnorm(x, mean = 1457, sd = 596)
x_sa_norm <- data.frame(cbind(x,y_sa))

x_sa_norm %>%
  ggplot(aes(x=x,y=y_sa)) +
  geom_line(color= "red")

```



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

co2_distr_sa <- data.frame(co2 = seq(400, 3000, .1)) %>%
  mutate(prob = dnorm(co2,1457,596))

sample_co2_sa <- sample(co2_distr_sa$co2, 1000, replace = TRUE, prob = co2_distr_sa$prob) #sample
sample_f_sa <- tibble(co2 = sample_co2_sa, f = ((co2-C_o)/C_a)/100000) %>%
  dplyr::select(-co2)
```

Quanta

I'll use the following studies for calculating the meanparameter:

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 Dhamadhakari (2012) : 17 patients, q: 138/34 (no mask/mask)

```
q <- (1.25*130+8.2*117+12.5+0.89*571+138*17)/(130+117+1+571+138) #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))
# reported quanta plus the two missing
q_sample_total_unif <- c(q_inf_persons, runif(117-length(q_inf_persons), min = 0, max =1))
```

```

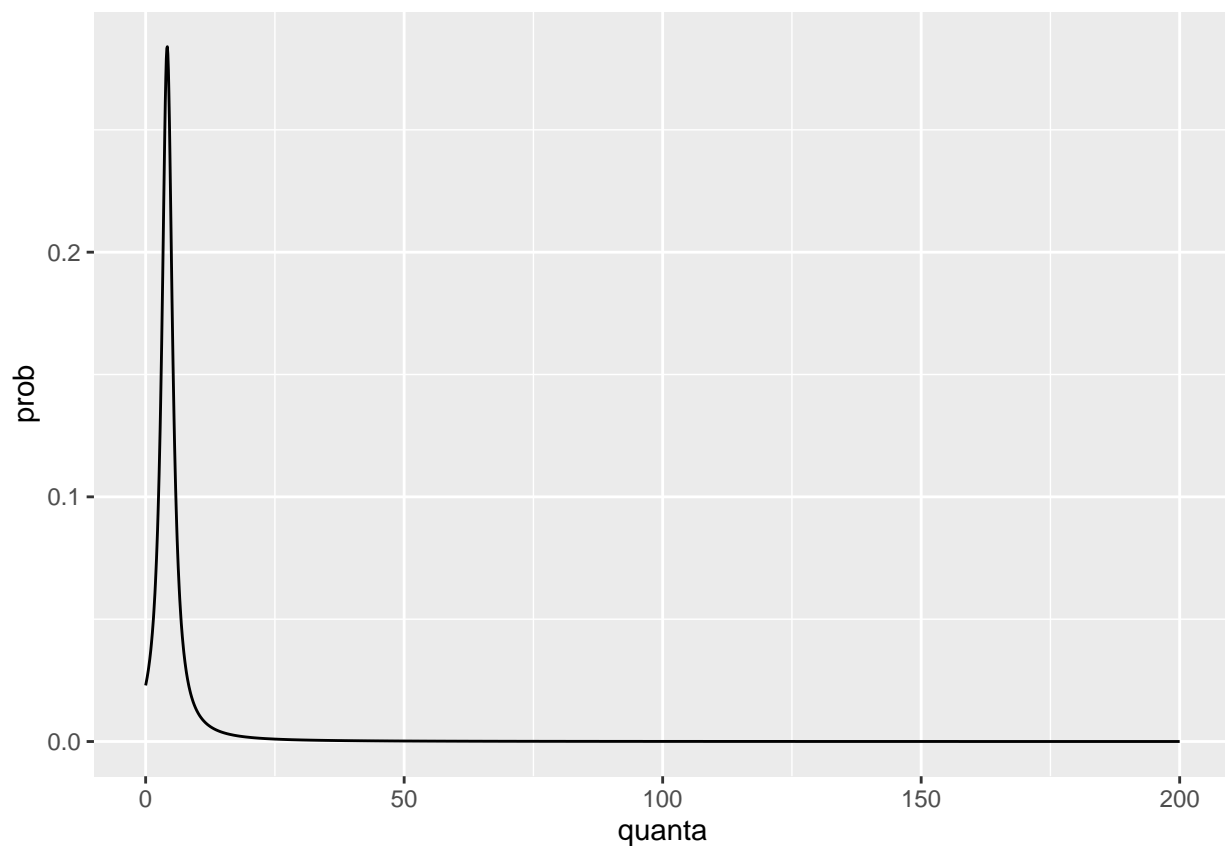
# rest unif in [0,1], as quanta below 1 isnt enough to infect an individual

### hier noch SD neu berechnen
dq <- function(x) {
  dtrunc(x, spec = "st", a = 0, b = 300, mu = q, sigma = 1.235, nu = 1) #sigma aus Escombe
} #function using parameters from option 1

rq_distr <- data.frame(quanta = seq(0, 200, .1)) %>%
  mutate(prob = dq(quanta))

plot2 <- ggplot(rq_distr, aes(x = quanta, y = prob)) +
  geom_line()
plot2

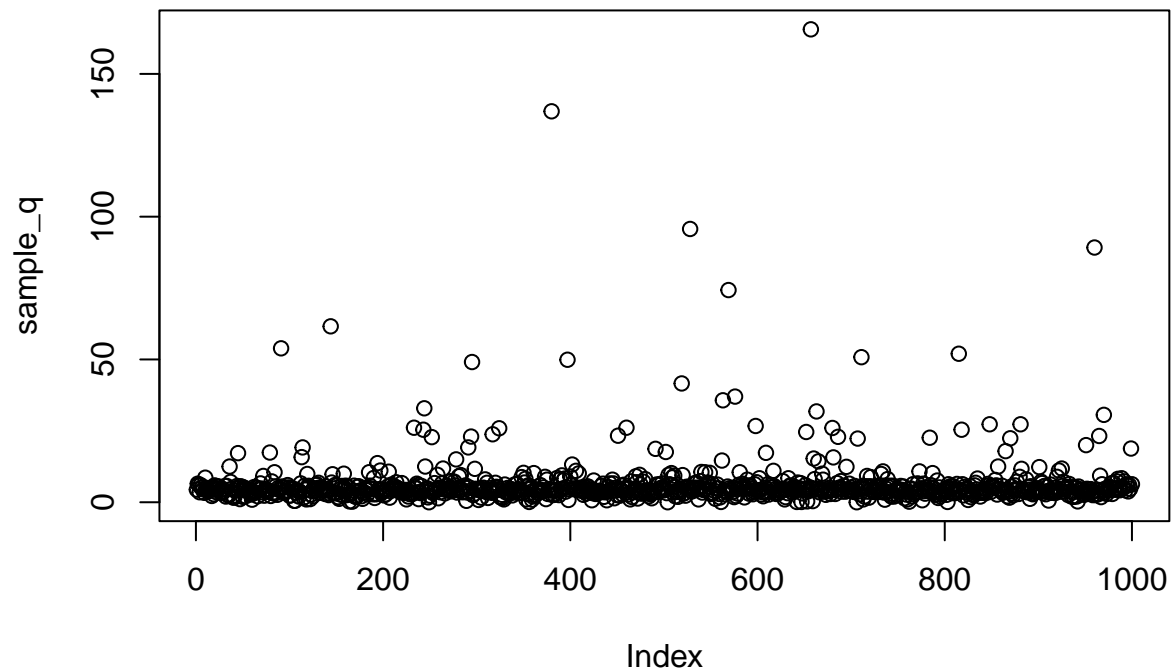
```



```

sample_q <- sample(rq_distr$quanta, 1000, replace = TRUE, prob = rq_distr$prob)
plot(sample_q)

```



Rest of the parameters

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

#I
prev_ch <- (0.46 + 8.23)/200000 #decimal; values for age group 10-14 and 15-19 over the last two years,
prev_sa <- 30000/59000000 #https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_Entity_type=%22cou
prev_tz <- 1177/61741120
#0.4 https://ntlp.go.tz/tuberculosis/paediatric-tb/

I_ch <- prev_ch*n_ch #prevalence per class (per year)
I_sa <- prev_sa*n_sa
I_tz <- prev_tz*n_sa

day <- 8
week <- 8*5
month <- 8*5*4
year <- 8*5*4*12
```

```

#preparing datasets for plotting
df_ch1 <- tibble(school = c(rep("school 1", 1000)), 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_week = 1 - exp(-(f*I_ch*q*week)/n_ch)) %>%
  mutate(P_day = 1 - exp(-(f*I_ch*q*day)/n_ch)) %>%
  mutate(P_year_one = 1 - exp(-(f*0.01*q*year)/n_ch)) %>%
  mutate(P_month_one = 1 - exp(-(f*0.01*q*month)/n_ch)) %>%
  mutate(P_week_one = 1 - exp(-(f*0.01*q*week)/n_ch)) %>%
  mutate(P_day_one = 1 - exp(-(f*0.01*q*day)/n_ch))

df_ch2 <- tibble(school = c(rep("school 2", 1000)), 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_week = 1 - exp(-(f*I_ch*q*week)/n_ch)) %>%
  mutate(P_day = 1 - exp(-(f*I_ch*q*day)/n_ch)) %>%
  mutate(P_year_one = 1 - exp(-(f*0.01*q*year)/n_ch)) %>%
  mutate(P_month_one = 1 - exp(-(f*0.01*q*month)/n_ch)) %>%
  mutate(P_week_one = 1 - exp(-(f*0.01*q*week)/n_ch)) %>%
  mutate(P_day_one = 1 - exp(-(f*0.01*q*day)/n_ch))

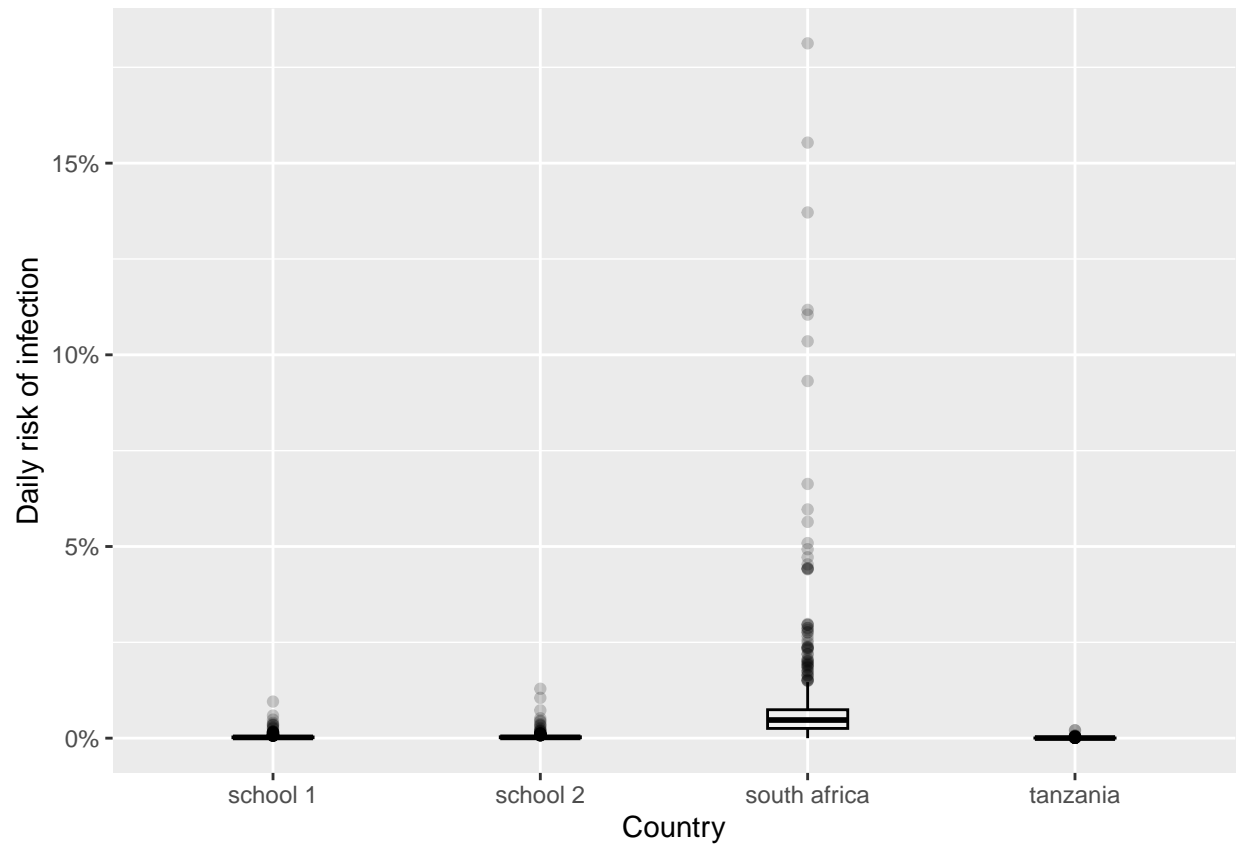
df_tz <- tibble(school = c(rep("tanzania", 1000)), f = sample_f_tz, q = sample_q) %>%
  mutate(P_year = 1 - exp(-(f*I_tz*q*year)/n_tz)) %>%
  mutate(P_month = 1 - exp(-(f*I_tz*q*month)/n_tz)) %>%
  mutate(P_week = 1 - exp(-(f*I_tz*q*week)/n_tz)) %>%
  mutate(P_day = 1 - exp(-(f*I_tz*q*day)/n_tz)) %>%
  mutate(P_year_one = 1 - exp(-(f*0.01*q*year)/n_ch)) %>%
  mutate(P_month_one = 1 - exp(-(f*0.01*q*month)/n_ch)) %>%
  mutate(P_week_one = 1 - exp(-(f*0.01*q*week)/n_ch)) %>%
  mutate(P_day_one = 1 - exp(-(f*0.01*q*day)/n_ch))

df_sa <- tibble(school = c(rep("south africa", 1000)), 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_week = 1 - exp(-(f*I_sa*q*week)/n_sa)) %>%
  mutate(P_day = 1 - exp(-(f*I_sa*q*day)/n_sa)) %>%
  mutate(P_year_one = 1 - exp(-(f*0.01*q*year)/n_ch)) %>%
  mutate(P_month_one = 1 - exp(-(f*0.01*q*month)/n_ch)) %>%
  mutate(P_week_one = 1 - exp(-(f*0.01*q*week)/n_ch)) %>%
  mutate(P_day_one = 1 - exp(-(f*0.01*q*day)/n_ch))

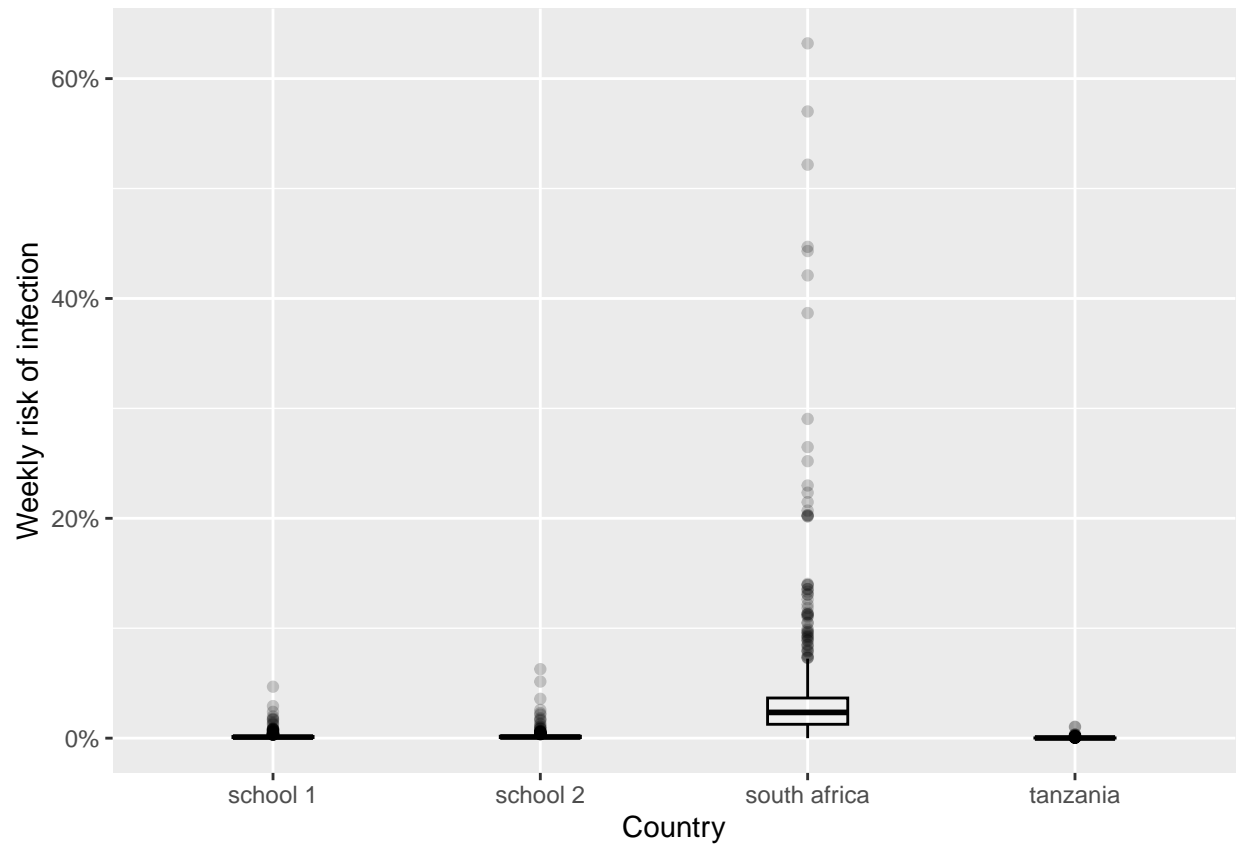
df_complet <- bind_rows(df_ch1, df_ch2, df_sa, df_tz)

df_complet %>%
  ggplot(aes(x = school, y=P_day$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("Daily risk of infection")

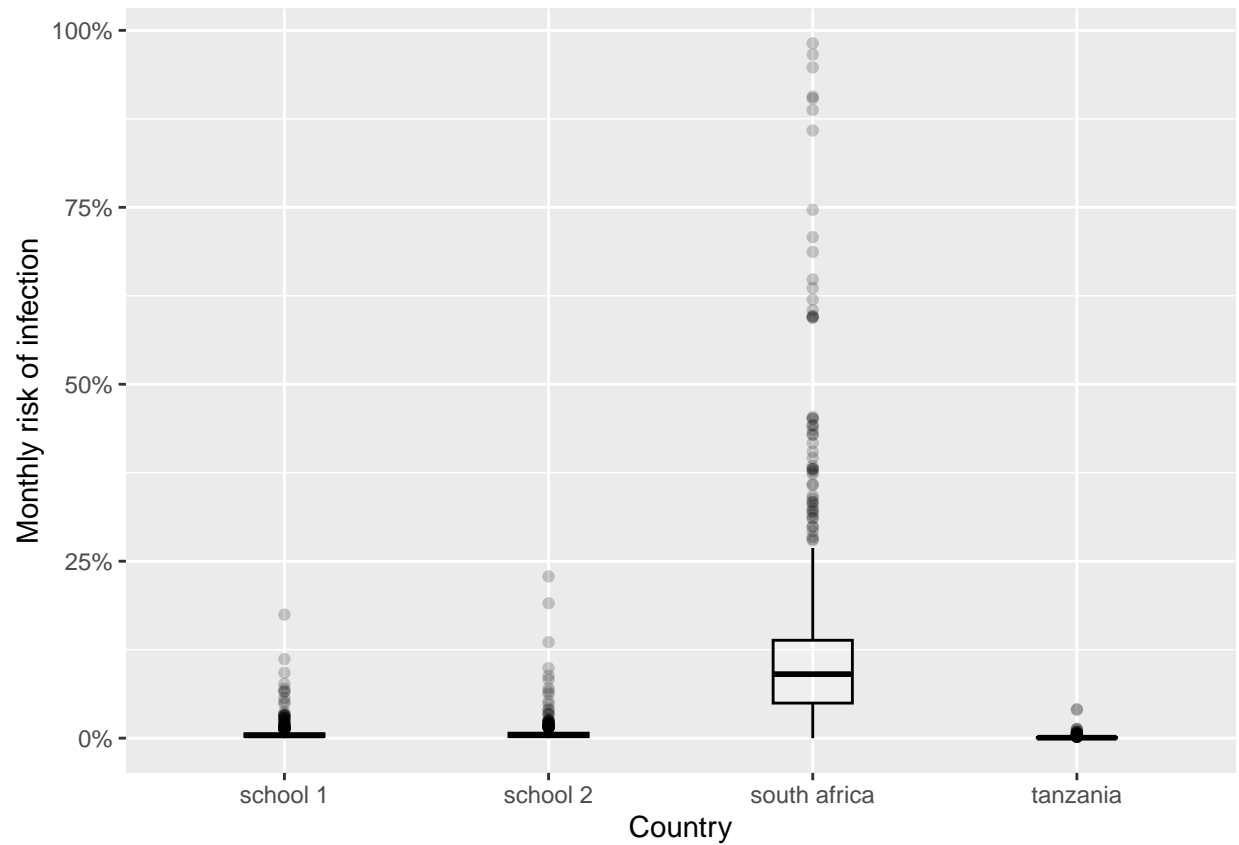
```



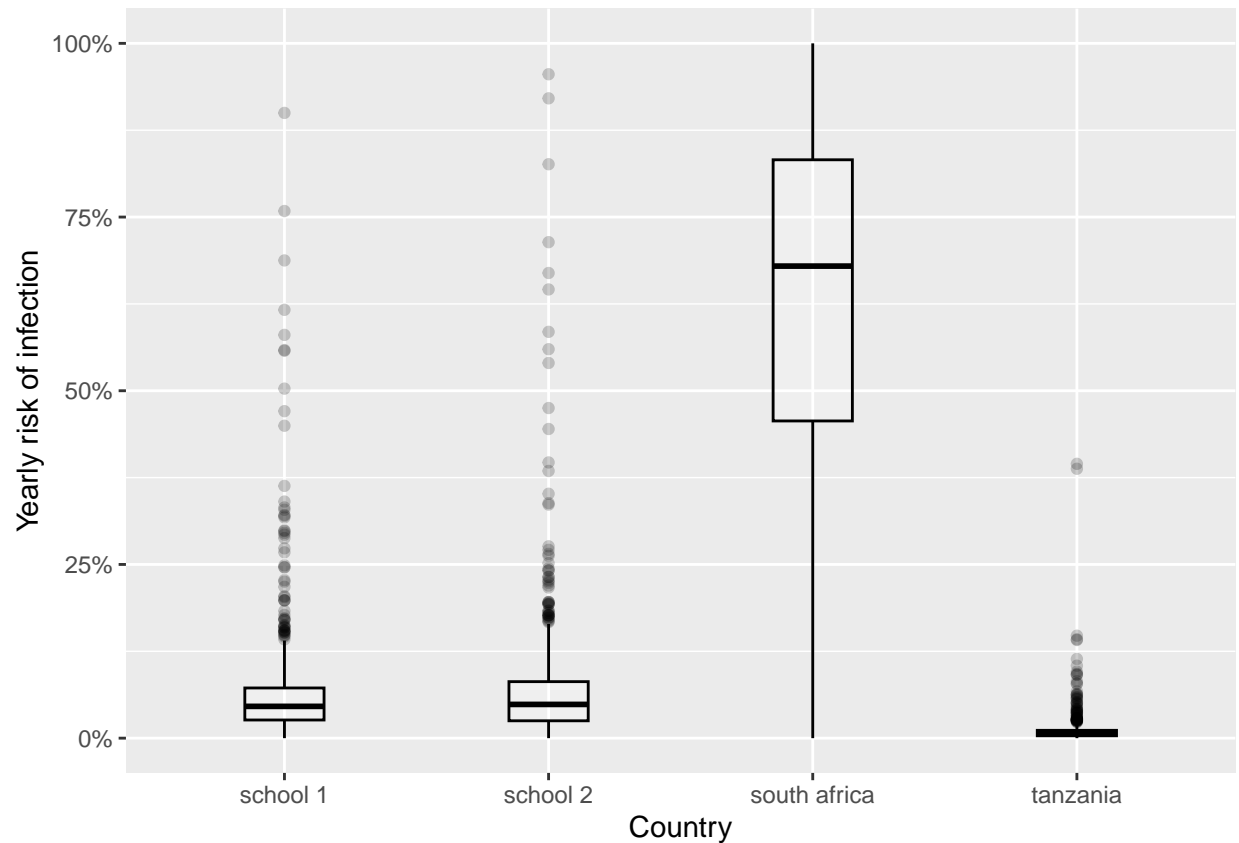
```
df_complet %>%
  ggplot(aes(x = school, y=P_week$f, colour = school))+
  geom_boxplot(width=0.3, color="black", alpha=0.2) +
  scale_y_continuous(labels = scales::percent_format(scale = 100)) +
  xlab("Country") +
  ylab("Weekly risk of infection")
```

```
df_complet %>%
  ggplot(aes(x = school, y=P_month$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("Monthly risk of infection")
```



```
df_complet %>%
  ggplot(aes(x = school, y=P_year$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")
```



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")
```

