```r
#Pre-work ----
library(readxl)
COVID_19 <- read_excel("C:/Users/82108/Downloads/COVID_19.xlsx")
View(COVID_19)
dim(COVID_19) #13418 column 11 row
library(dplyr)
filter(COVID_19, year == 2019)
arrange(filter(COVID_19, year == 2019))
arrange(filter(COVID_19, month == 3 & continentExp == "Asia"), deaths)

#ggplot workflow ----
##data cleaning ----
rm(list=ls())
library(ggplot2)
library(readxl)
COVID_19 <- read_excel("C:/Users/82108/Downloads/COVID_19.xlsx")
View(COVID_19)
dim(COVID_19) #13418 column 11 row
library(dplyr)
SPR_US <- data.frame (arrange(filter(COVID_19, countryterritoryCode == "USA"
                          & month == 3), day))
SPR_US %>% glimpse
SPR_US %>% dim #31 column 11 row
SPR_US

##data extraction & labelling ----
Date.label <- "dateRep"
Date <- SPR_US[, Date.label]
Date %>% class #Output: "POSIXct" "POSIxt"_ is this a factor? numerical?

Cases.label <- "cases"
Cases <- SPR_US[, Cases.label]
Cases %>% class #Output: "numeric"

Deaths.label <- "deaths"
Deaths <- SPR_US[, Deaths.label]
Deaths %>% class #Output: "numeric"

##Plot 1.0: dateRep, cases ----
plot.base.cases <- ggplot(SPR_US, aes(x= Date, y= Cases))+
  ggtitle("# of cases over time")+
  xlab(Date.label)+
  ylab(Cases.label)+
  theme_minimal()+
  theme(plot.title = element_text(size = 16, hjust = 0.5))
plot.base.cases

##Plot 2.0: dateRep, deaths ----
plot.base.deaths <- ggplot(SPR_US, aes(x= Date, y= Deaths))+
  ggtitle("# of deaths over time")+
  xlab(Date.label)+
  ylab(Deaths.label)+
  theme_minimal()+
  theme(plot.title = element_text(size = 16, hjust = 0.5))
plot.base.deaths

##Boxplot 1.1 ----
plot.base.cases +
  geom_boxplot(fill = "lightgrey", color = "black")

##Scatterplot 1.1 ----
plot.base.cases +
  geom_point(aes(color = Deaths), position = "jitter", size = 3)

##Scatterplot 2.1 ----
plot.base.deaths +
  geom_point() +
  geom_text(aes(label = Deaths.label), vjust= 1.5, colour = "darkgrey",
            position = position_dodge(.9), size = 3)

##Scatterplot 2.2 ----
plot.base.deaths +
  geom_point(position = "jitter") +
```

```
  geom_hline(yintercept = mean(Deaths), colour = "blue")

library(rmarkdown)
library(latexpdf)
library(tinytex)
```