

# STA442 Assignment 3

Maggie Ma | 1004134858

18/11/2020

## Question 1 - CO2 Data

### Introduction

The research question of interest involves the effect of major industrial and economic human activities on the concentration level of atmospheric Carbon Dioxide ( $CO_2$ ), as observed at Mauna Loa Observatory in Hawaii. More specifically, we are interested in looking at two specific events: the fall of the Berlin wall in November 1989 and the global lockdown during the COVID-19 pandemic starting in February 2020.

### Methods

In order to fit a statistically robust model that quantifies the problem, data collected at Mauna Loa Observatory in Hawaii, during the time period of January 1960 - June 2020, were used to analyze the students math scores. This data was made available by the Scripps  $CO_2$  Program at scrippsco2.ucsd.edu.

Since the gamma family is often used to model positive and continuous data and has nice properties when working with prior and posterior distributions, we fit a generalized additive model with gamma distribution to answer the research question: whether the fall of the Berlin wall in November 1989 and the global lockdown during the COVID-19 pandemic starting in February 2020 each has a effect on the concentration level of  $CO_2$ . The model is shown below:

$$\begin{aligned}Y_i &\sim \text{Gamma}(\lambda(t_i)) \\ \log(\lambda(t)) &= X_i\beta + s(t) + U_i \\ U_i &\sim \text{RW2}(\sigma^2) \\ X_{i0} &= 1 \\ X_{i1} &= \cos(2\pi t_i/365.25) \\ X_{i2} &= \sin(2\pi t_i/365.25) \\ X_{i3} &= \cos(2\pi t_i/182.625) \\ X_{i4} &= \sin(2\pi t_i/182.625)\end{aligned}$$

In this model, the concentration level of  $CO_2$   $Y_i$  follows a gamma distribution and is a function of  $\lambda_t$ . In order to capture the seasonal variation of  $CO_2$ , four sinusoidal basis functions  $X_i$ s were used to capture a wide range of seasonal effects using 12 month and a 6 month frequency cycles.  $s(t)$  is the sum of these 4 sinusoids and is a smoothing parameter. Lastly,  $U_i$  is the random effect of time following a random slope,  $\sigma$  is in rate per year, with an exponential prior  $param = c(0.1, 0.5)$  indicating a median of 0.1. It is a small number because we anticipate the slope change from one year to the next by 0.1.

### Results

Table 1: Standard Deviation from GAM model results

	0.5quant	0.025quant	0.975quant
sd for gamma	0.00009	0.000089	0.000092
sd for timeInla	0.00211	0.002097	0.002118

Table 1 represents the standard deviation for gamma and time. Standard deviation for gamma is very small, indicating we don't have a big need for gamma error, the observations on  $CO_2$  doesn't look very noisy as anticipated, this result is also shown in Figure 1, where confidence interval increases for predicted periods after June 2020. We can also see a clear seasonal variation in  $CO_2$  level. Standard deviation for timeInla looks at changes in slope over a one year period, similar to our estimation for prior, we are not surprised to see that the slope doesn't change very much in a one year period.

Figure 1.: Posterior Mean of Fitted Values

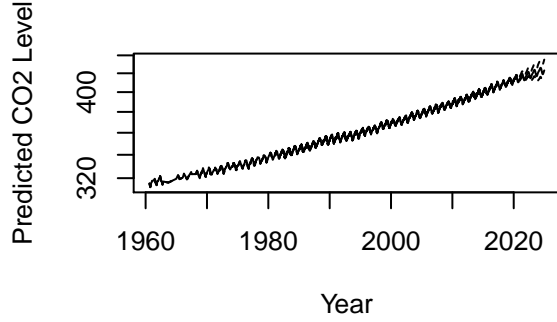


Figure 2.: Posterior Sample Means for Random Effect

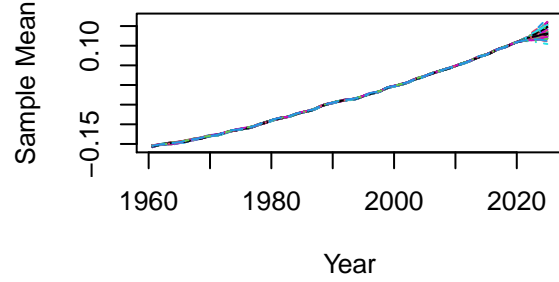


Figure 3.: Derivative of Posterior Sample Means for Random Effect  
Jan 1960 – Dec 2020

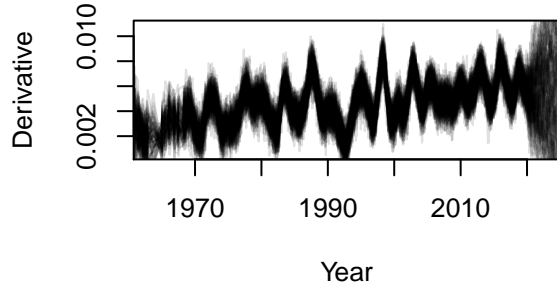
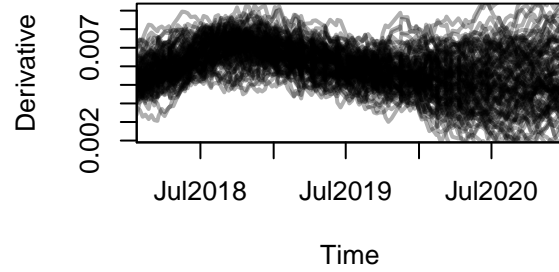


Figure 4.: Derivative of Posterior Sample Means for Random Effect,  
Jan 2018 – Dec 2020



In order to examine the effect of the two events, we took 100 posterior random samples from the random effect and created visualizations for posterior sample means for random effect and the derivatives of these means (Figure 2, 3, 4). The fall of Berlin wall that caused a dramatic fall in industrial production in the Soviet Union and Eastern Europe occurred in 1989, from the posterior sample means for random effect (figure 2), we can see that the years following 1989 has a relatively flat curve comparing to the other years, concluding a significant effect on  $CO_2$  concentration level. We cannot really infer anything about the COVID-19 lockdown, however, from this graph because random samples are all over the place past June 2020. Figure 3 and 4 are the derivative of the posterior sample means, figure 4 focuses on a recent time period that allows us to further investigate. All derivatives are positive, indicating that  $CO_2$  levels are on the rise. Unfortunately, the 100 random samples taken do not show a clear pattern of where the predicted  $CO_2$  concentration level is headed post Jun 2020.

## Conclusion

From a range of visual investigations, we conclude that the fall of the Berlin wall in November 1989, which preceded with a dramatic fall in industrial production in the Soviet Union and Eastern Europe have had a lasting effect for the years to follow on the concentration level of  $CO_2$ . The increase in  $CO_2$  has slowed down

because of this event. As for the effect of global lockdown during COVID-19, our actual data of  $CO_2$  stops at June 2020, and we cannot conclude where the predicted level of  $CO_2$  is headed. Therefore, we do not know whether the lockdown during COVID-19 in 2020 has an effect on the concentration level of  $CO_2$ .

## Question 2 - Quebec Mortality Data

### Introduction

The COVID-19 global pandemic reached Canada in early March. Provinces across Canada have all been impacted in various ways, with Ontario and Quebec having the most number of people infected. This research focuses on the daily mortality counts in Quebec as it has only been made available recently. The research question of interest involves how COVID infected different age groups at various time periods. It is hypothesized that during the first wave in March, April, and May, the epidemic primarily affected the elderly, in particular over age of 70, as seen from the weekly mortality counts, not only because they are one of the most vulnerable groups to COVID, but nursing home was also one of the hardest hit places where many of them reside. However, our second part of the hypothesis states that the second wave is caused by certain young people acting irresponsibly, as there is an increase in deaths in the under 50s in most recent death data whereas the over 70's have no more deaths than would be expected pre-covid.

### Methods

In order to fit a statistically robust model that quantifies the problem, mortality data recently released from the government of Quebec was used to analyze the excess deaths in different age groups at various time periods of the pandemic. Since the Poisson family is used for positive and discrete data, we used Bayesian inference with semi-parametric smoothed time trend to fit a Poisson model to the data. We first used the data prior to March 1st as pre-COVID data to predict mortality what would have otherwise been without COVID in the recent months taking seasonal variations into consideration. We then took 100 random posteriors samples and calculated excess deaths by comparing predicted mortality numbers with actual numbers. This model was ran once for deaths of people aged 70 and above, and once for people aged 50 and under, in order to examine the hypothesis that the first wave of COVID-19 primary affected the elderly (70+) and second wave primarily affected younger people (<50). The model is shown below:

$$\begin{aligned} Y_i &\sim \text{Poisson}(O_i \lambda_i) \\ \log(\lambda_i) &= X_i \beta + U_i + V_i \\ U_i &\sim \text{RW2}(\sigma^2) \\ V_i &\sim N(0, \tau^2) \\ X_{i1} &= \cos(2\pi t_i / 365.25) \\ X_{i2} &= \sin(2\pi t_i / 365.25) \\ X_{i3} &= \cos(2\pi t_i / 182.625) \\ X_{i4} &= \sin(2\pi t_i / 182.625) \end{aligned}$$

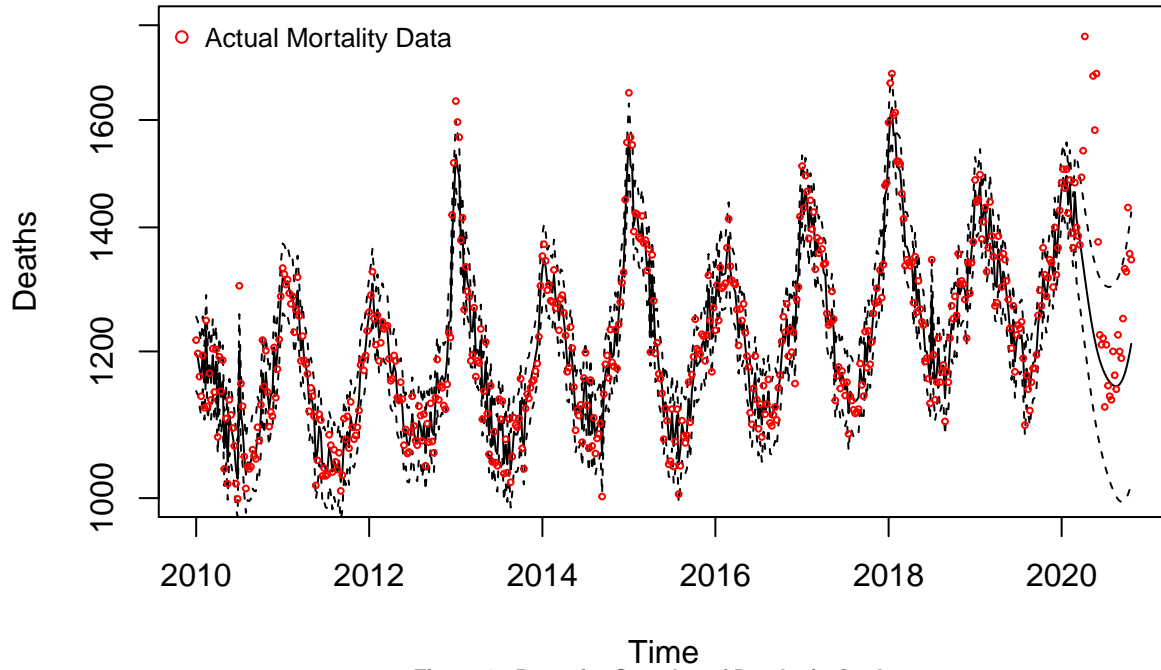
In this model, the number of deaths  $Y_i$  follows a Poisson distribution. In order to capture the seasonal variation of deaths, four sinusoidal basis functions  $X_i$ s were used to capture a wide range of seasonal effects using 12 month and a 6 month frequency cycles.  $U_i$  follows the second random walk - random slope, and answers the question 'how much does slope change from one year to the next'.  $\sigma$  is in rate per year, with an exponential prior  $\text{param} = c(0.1, 0.5)$  indicating a median of 0.01. It is a small number because we anticipate the slope change from one year to the next by 0.01.  $V_i$  is the independent time variation for over-dispersion and follows a exponential prior of  $\text{param} = c(\log(1.2), 0.5)$ , indicating a prior median  $\tau$  of  $\log(1.2)$ . This prior was chosen based on our assumption that there will be 20% more deaths and with COVID pandemic, this is a reasonable guess.

### Results

When examining COVID related questions, it is important first to recognize the severity of COVID and that COVID-19 is indeed more deadly than the common flu. Figure 1 shows the predicted number of deaths in a world without COVID and the actual deaths during COVID months. Similarly, figure 2 represents deaths generated from 100 posterior samples in a COVID-free world, overlayed with actual data on top. March 1st was determined as the day COVID hit Canada, therefore the cut off point for data used for predictions. It is easily observed from the figure that excess death was created as a result of COVID, especially around April (first wave) and recent months (second wave). Indeed, the median of excess deaths from posterior samples

suggest that the total number of excess death, most likely caused by COVID, from March 1st to October 21st is around 6000, as suggested in Figure 3.

**Figure 1.: Predicted and Actual Number of Deaths in Quebec**



**Figure 2.: Posterior Samples of Deaths in Quebec**

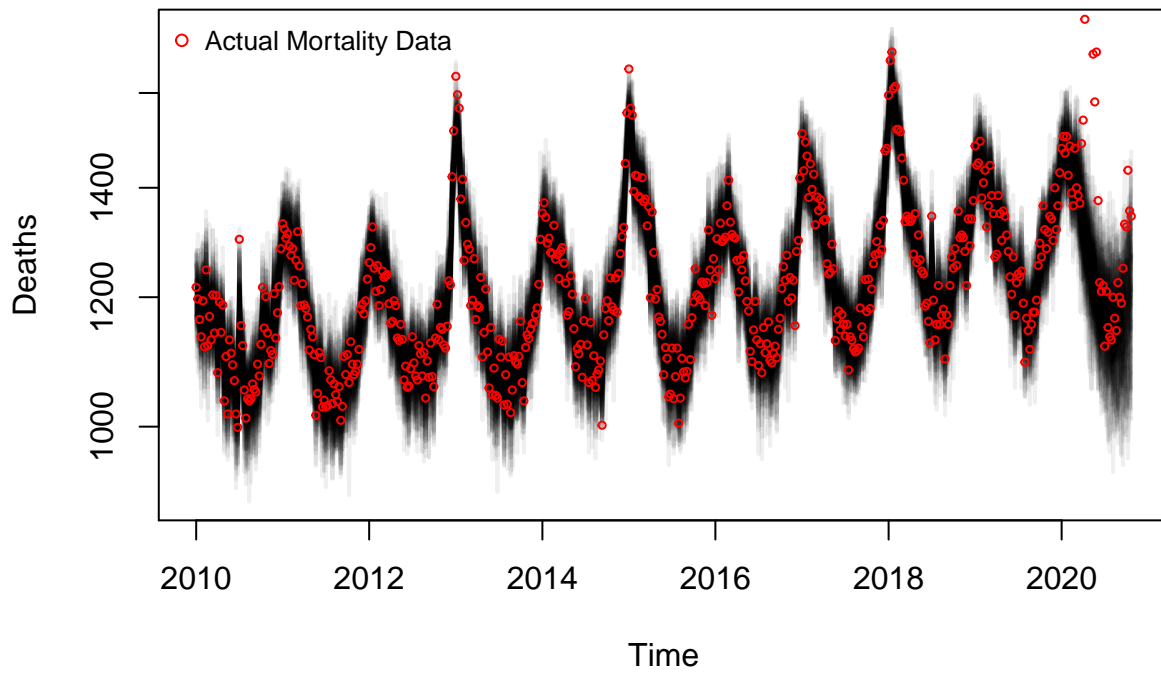
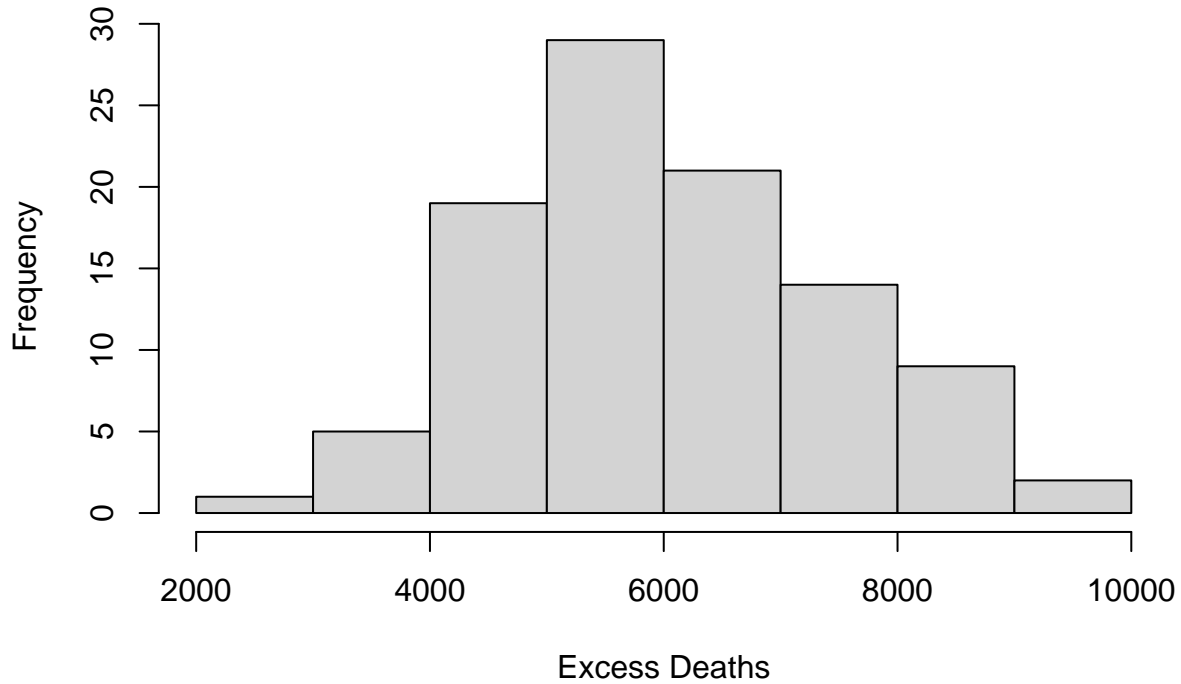


Table 2: Quantile of Posterior Sample Excess Deaths, March 1st - October 21st, 2020

Excess Deaths	
0%	2822
25%	5015
50%	5924
75%	7004
100%	9759

Figure 3.: Histogram of Total Excess Deaths from Posterior Samples, March 1st – October 21st, 2020



The first part of our hypothesis is that the first wave of COVID, from March to May, primarily affected older people. Table 3 shows the quantile of posterior sample excess deaths, from March 1st to May 31st, for the age group 70 and above and all age groups. Excess deaths for all population is around 4700 deaths for this time period, while excess deaths for 70+ age group is around 4400, making up 94% of all deaths. This confirms our first hypothesis that first wave of COVID affected primarily older people.

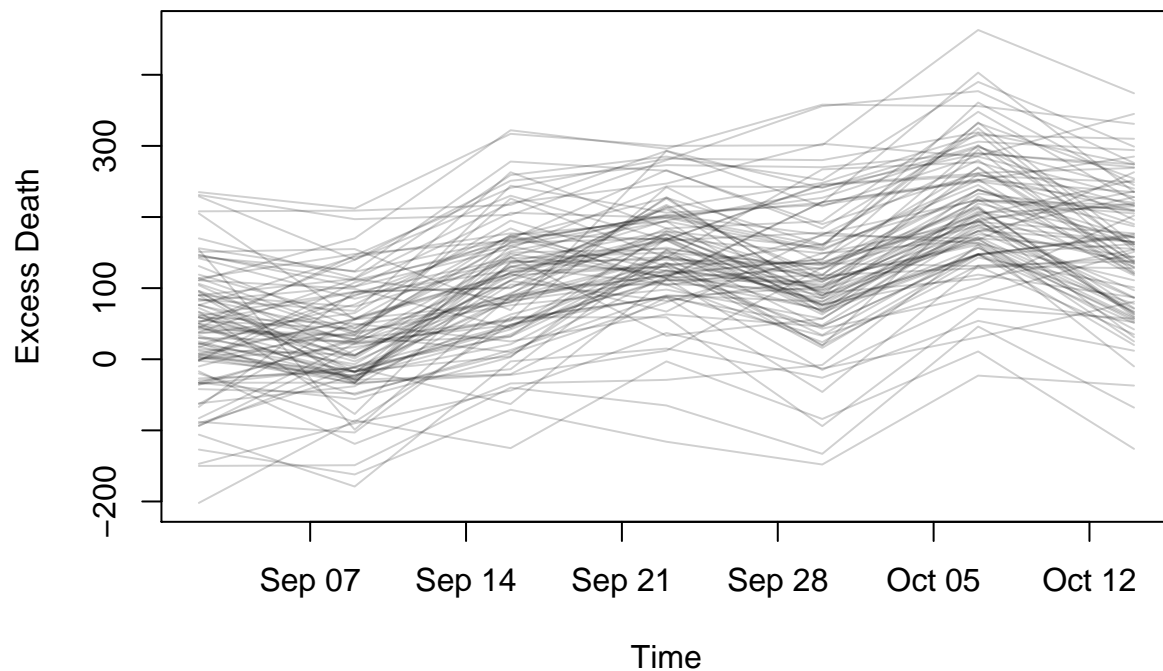
Table 3: Quantiles of Posterior Sample Excess Deaths, March 1st - May 31st, 2020

	70+ Years of Age	Total
0%	3379	3889
25%	4066	4524
50%	4354	4798
75%	4708	5079
100%	5234	5793

As for our second hypothesis, we calculated excess deaths among 70+ year old age group (as shown in figure 4) and among 0-49 year old age group (as shown in figure 5) during both first and second wave of COVID-19.

We define first wave as the time between March 1st and May 31st. We define the second wave as the time between September 1st and the most recent data, during the week of October 21st. As seen from table 4, the older population is again, similarly to the first wave, most vulnerable to COVID-19, as the excess deaths is around 110, whereas the excess death for 0-49 year olds is around 12.

**Figure 4.: Posterior Sample Excess Deaths Among 70+ yo**



**Figure 5.: Posterior Sample Excess Deaths Among 0-49 yo**

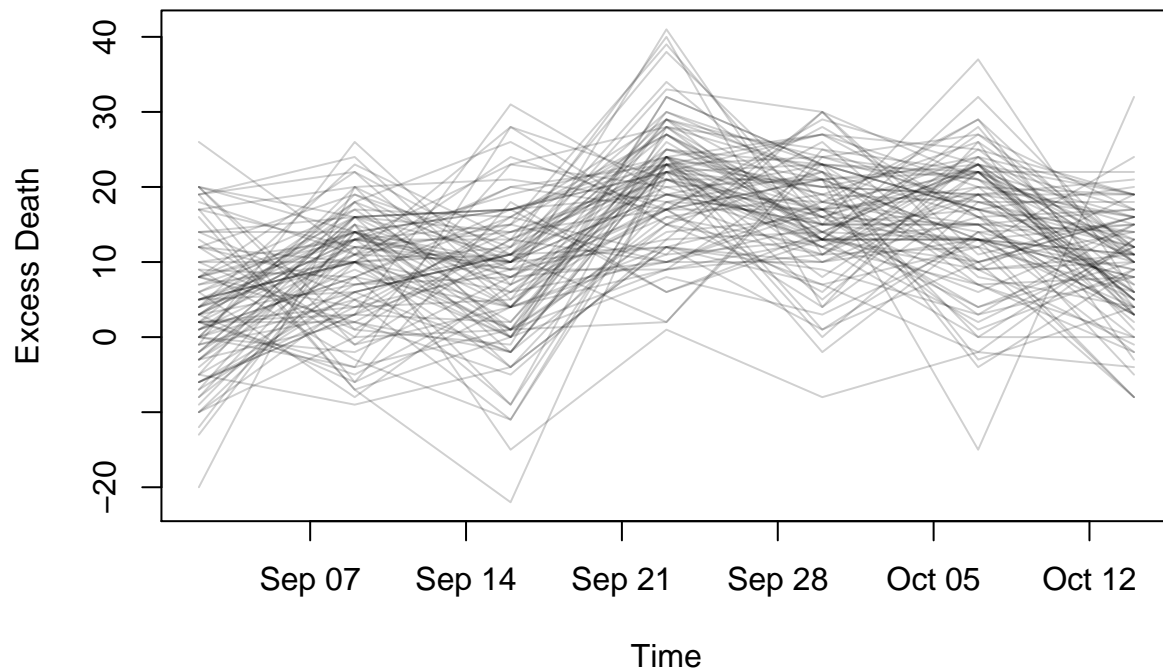


Table 4: Quantile of Posterior Sample Excess Deaths, Sep 1st - Oct 21st, 2020

	70+ Year of Age	0-49 Years of Age
0%	-202	-22
25%	46	5
50%	118	12
75%	184	19
100%	463	41

### Discussion

COVID-19 is more serious than the seasonal flu and has caused great number of excess deaths. Although many may believe the second wave was caused by irresponsible young people, we conclude that young people are not the most vulnerable population from such infectious disease. In fact, excess deaths are highest among people aged 70 and above, during both first and second wave of the pandemic. Further investigation is needed on infection data in order to draw conclusion on whether it was mostly young people who were involved in the spreading of COVID-19 during second wave.



## Appendix

### Question 1 Codes

```
# get data
cUrl = paste0("http://scrippsco2.ucsd.edu/assets/data/atmospheric/", "stations/flask_co2/daily/daily_fl")
cFile = basename(cUrl)
if (!file.exists(cFile)) download.file(cUrl, cFile)
co2s = read.table(cFile, header = FALSE, sep = ",", skip = 69, stringsAsFactors
                 = FALSE, col.names = c("day", "time", "junk1", "junk2",
                                         "Nflasks", "quality", "co2"))
co2s$date = strptime(paste(co2s$day, co2s$time), format = "%Y-%m-%d %H:%M",
                    tz = "UTC")
# remove low-quality measurements
co2s = co2s[co2s$quality == 0, ]
# plot
plot(co2s$date, co2s$co2, log = "y", cex = 0.3, col = "#00000040",
     xlab = "time", ylab = "ppm")
plot(co2s[co2s$date > ISOdate(2015, 3, 1, tz = "UTC"), c("date", "co2")],
     log = "y", type = "o", xlab = "time", ylab = "ppm", cex = 0.5)

# create time variable for inla
co2s$day = as.Date(co2s$date)
# adding extra data for forecast
## start 3 days so we don't get any duplicates
toAdd = data.frame(day = seq(max(co2s$day) + 3, as.Date("2025/1/1"),
                             by = "10 days"), co2 = NA)
co2ext = rbind(co2s[, colnames(toAdd)], toAdd)
## make time origin 2000 instead of 1970
timeOrigin = as.Date("2000/1/1")
## create years
co2ext$timeInla = round(as.numeric(co2ext$day - timeOrigin)/365.25, 2)

# creating sin and cos as basis functions
co2ext$cos12 = cos(2*pi*co2ext$timeInla) #time multiplied by 2 pi
co2ext$sin12 = sin(2*pi*co2ext$timeInla)
co2ext$cos6 = cos(2*2*pi*co2ext$timeInla)
co2ext$sin6 = sin(2*2*pi*co2ext$timeInla)

# disable some error checking in INLA
mm = get("inla.models", INLA:::inla.get.inlaEnv())
if(class(mm) == 'function') mm = mm()
mm$latent$rw2$min.diff = NULL
assign("inla.models", mm, INLA:::inla.get.inlaEnv())

co2res = inla(co2 ~ sin12 + cos12 + sin6 + cos6 +
              f(timeInla, model = 'rw2',
                prior='pc.prec', param = c(0.1, 0.5)),
              data = co2ext, family='gamma',
              control.family = list(hyper=list(
                prec=list(prior='pc.prec', param=c(0.1, 0.5)))),
              control.predictor = list(compute=TRUE, link=1),
```

```

control.compute = list(config=TRUE), verbose=FALSE)

qCols = c('0.5quant','0.025quant','0.975quant')
Pmisc::priorPost(co2res)$summary[,qCols]

# inla posterior sample function
sampleList = INLA::inla.posterior.sample(100, co2res,
                                         selection = list(timeInla = 0))
sampleMean = do.call(cbind, Biobase::subListExtract(sampleList, "latent"))
sampleDeriv = apply(sampleMean, 2, diff)/diff(co2res$summary.random$timeInla$ID)
# plot 1 posterior mean of fitted values
### fitted values stop in 2020, CI start to get wider until 2025
matplot(co2ext$day, co2res$summary.fitted.values[, qCols], type = "l",
        col = "black", lty = c(1, 2, 2), log = "y", xlab = "time", ylab = "ppm")
# plot 2 random effect f()
## flatten out in 1989
Stime = timeOrigin + round(365.25 * co2res$summary.random$timeInla$ID)
matplot(Stime, co2res$summary.random$timeInla[, qCols], type = "l",
        col = "black", lty = c(1, 2, 2), xlab = "time", ylab = "y")

# extra plot: posterior sample means for random effect
## 1989 is kinda flat, it 2020 flat? not sure
matplot(Stime, sampleMean, type = 'l')

# plot 3 derivative
### derivative is positive everywhere, so the CO2 is definitely going up except
## maybe 1992, 2020 is hard to say
matplot(Stime[-1], sampleDeriv, type = "l", lty = 1, xaxs = "i",
        col = "#00000020",
        xlab = "time", ylab = "deriv",
        ylim = quantile(sampleDeriv, c(0.01, 0.995)))

# plot 4 derivative, but only 2018-2021
# don't know but maybe it'll take a while to show, actual data only till june 2020
forX = as.Date(c("2018/1/1", "2021/1/1"))
forX = seq(forX[1], forX[2], by = "6 months")
toPlot = which(Stime > min(forX) & Stime < max(forX))
matplot(Stime[toPlot], sampleDeriv[toPlot, ], type = "l",
        lty = 1, lwd = 2, xaxs = "i", col = "#00000050",
        xlab = "time", ylab = "deriv", xaxt = "n", ylim = quantile(sampleDeriv[toPlot,
], c(0.01, 0.995)))
axis(1, as.numeric(forX), format(forX, "%b%Y"))

```

## Question 2 Codes

```

# download data
xWide = read.table(paste0("https://www.stat.gouv.qc.ca/statistiques/",
                          "population-demographie/deces-mortalite/",
                          "WeeklyDeaths_QC_2010-2020_AgeGr.csv"),
                  sep = ";", skip = 7,
                  col.names = c("year", "junk", "age", paste0("w", 1:53)))
xWide = xWide[grep("^[[[:digit:]]+$", xWide$year), ]
x = reshape2::melt(xWide, id.vars = c("year", "age"),

```

```

        measure.vars = grep("~w[[:digit:]]+$",
                             colnames(xWide)))
x$dead = as.numeric(gsub("[[:space:]]", "", x$value))
x$week = as.numeric(gsub("w", "", x$variable))
x$year = as.numeric(x$year)
x = x[order(x$year, x$week, x$age), ]

#convert the 'week' variable to time
newYearsDay = as.Date(ISOdate(x$year, 1, 1))
x$time = newYearsDay + 7 * (x$week - 1)
x = x[!is.na(x$dead), ]
x = x[x$week < 53, ]

#Divide the data into pre and post covid, add extra dates to data
#so that INLA will create forecasts.
dateCutoff = as.Date("2020/3/1")
xPreCovid = x[x$time < dateCutoff, ]
xPostCovid = x[x$time >= dateCutoff, ]
toForecast = expand.grid(age = unique(x$age), time = unique(xPostCovid$time),
                        dead = NA)
xForInla = rbind(xPreCovid[, colnames(toForecast)], toForecast)
xForInla = xForInla[order(xForInla$time, xForInla$age),]
#Create some time variables, including sines and cosines. Time in years and
#centred is numerically stable in INLA.
xForInla$timeNumeric = as.numeric(xForInla$time)
xForInla$timeForInla = (xForInla$timeNumeric -
                        as.numeric(as.Date("2015/1/1")))/365.25
xForInla$timeId = xForInla$timeNumeric
xForInla$sin12 = sin(2 * pi * xForInla$timeNumeric/365.25)
xForInla$sin6 = sin(2 * pi * xForInla$timeNumeric *2/365.25)
xForInla$cos12 = cos(2 * pi * xForInla$timeNumeric/365.25)
xForInla$cos6 = cos(2 * pi * xForInla$timeNumeric *2/365.25)

# fit a model for total deaths in INLA
xForInlaTotal= xForInla[xForInla$age == 'Total', ]
##### MODEL #####
res = inla(dead ~ sin12 + sin6 + cos12 + cos6 +
           f(timeId, prior='pc.prec', param= c(log(1.2), 0.5)) +
           f(timeForInla, model = 'rw2', prior='pc.prec', param= c(0.01, 0.5)),
           data=xForInlaTotal,
           control.predictor = list(compute=TRUE, link=1),
           control.compute = list(config=TRUE), family='poisson')

### TOTAL
par(mfrow = c(1, 2))
# Plot predicted intensity and random effect
matplot(xForInlaTotal$time, res$summary.fitted.values[, qCols], type= "l",
        ylim = c(1000, 1800), lty = c(1, 2, 2), col = "black", log = "y",
        main = 'Figure 1.: Predicted and Actual Number of Deaths in Quebec',
        ylab = 'Deaths',
        xlab = 'Time', cex.main = 0.7)
points(x[x$age == "Total", c("time", "dead")], cex = 0.4, col = "red")
legend("topleft", col = 'red', legend = 'Actual Mortality Data', bty = "n",

```

```

    pch = 1, cex = 0.8)
# Take posterior samples of the intensity
sampleList = INLA::inla.posterior.sample(100, res,
                                         selection = list(Predictor = 0))
sampleIntensity = exp(do.call(cbind,
                              Biobase::subListExtract(sampleList, "latent")))
sampleDeaths = matrix(rpois(length(sampleIntensity),
                              sampleIntensity), nrow(sampleIntensity),
                      ncol(sampleIntensity))
# plot samples and real data
matplot(xForInlaTotal$time, sampleDeaths, col = "#00000010",
        lwd = 2, lty = 1, type = "l", log = "y",
        main = 'Figure 2.: Posterior Samples of Deaths in Quebec',
        ylab = 'Deaths',
        xlab = 'Time', cex.main = 0.7)
points(x[x$age == "Total", c("time", "dead")], col = "red", cex = 0.5)
legend("topleft", col = 'red', legend = 'Actual Mortality Data', bty = "n",
      pch = 1, cex = 0.8)

# calculate excess deaths
xPostCovidTotal = xPostCovid[xPostCovid$age == "Total", ]
xPostCovidForecast = sampleDeaths[match(xPostCovidTotal$time,
                                         xForInlaTotal$time), ]
excessDeaths = xPostCovidTotal$dead - xPostCovidForecast

# Total excess deaths march-may inclusive
par(mfrow = c(1, 1))
excessDeathsSub = excessDeaths[xPostCovidTotal$time > as.Date("2020/03/01")
                               & xPostCovidTotal$time < as.Date("2020/10/21"), ]
excessDeathsInPeriod = apply(excessDeathsSub, 2, sum)
b<- round(quantile(excessDeathsInPeriod))
table2 <- b %>%
  knitr::kable(caption = "Quantile of Posterior Sample Excess Deaths,
                        March 1st - October 21st, 2020" , booktabs = T, digits = 6,
              col.names = 'Excess Deaths') %>%
  kableExtra::kable_classic(full_width = F, html_font = "Cambria") %>%
  kableExtra::kable_styling(latex_options = c("HOLD_position", "striped"))
table2
hist(excessDeathsInPeriod, main = 'Figure 3.: Histogram of Total Excess
  Deaths from Posterior Samples, March 1st - October 21st, 2020',
  xlab = 'Excess Deaths', cex.main = 0.7)

### HYPOTHESIS 1
# total
# calculate excess deaths
xPostCovidTotal = xPostCovid[xPostCovid$age == "Total", ]
xPostCovidForecast = sampleDeaths[match(xPostCovidTotal$time,
                                         xForInlaTotal$time), ]
excessDeaths = xPostCovidTotal$dead - xPostCovidForecast
# Total excess deaths march-may inclusive
excessDeathsSub = excessDeaths[xPostCovidTotal$time > as.Date("2020/03/01")
                               & xPostCovidTotal$time < as.Date("2020/06/01"), ]

```

```

excessDeathsInPeriod = apply(excessDeathsSub, 2, sum)
wave1_total <- round(quantile(excessDeathsInPeriod))

# 70+
# fit a model for 70+ yo deaths in INLA
xForInlaTotal70= xForInla[xForInla$age == '70 years old and over', ]
# model
res70 = inla(dead ~ sin12 + sin6 + cos12 + cos6 +
  f(timeId, prior='pc.prec', param= c(log(1.2), 0.5)) +
  f(timeForInla, model = 'rw2', prior='pc.prec',
    param= c(0.01, 0.5)),
  data=xForInlaTotal70,
  control.predictor = list(compute=TRUE, link=1),
  control.compute = list(config=TRUE), family='poisson')
# Take posterior samples of the intensity
sampleList = INLA::inla.posterior.sample(100, res70,
  selection = list(Predictor = 0))
sampleIntensity = exp(do.call(cbind,
  Biobase::subListExtract(sampleList, "latent")))
sampleDeaths = matrix(rpois(length(sampleIntensity),
  sampleIntensity), nrow(sampleIntensity),
  ncol(sampleIntensity))

## 70+ excess death
# calculate excess deaths
xPostCovidTotal = xPostCovid[xPostCovid$age == "70 years old and over", ]
xPostCovidForecast = sampleDeaths[match(xPostCovidTotal$time,
  xForInlaTotal$time), ]
excessDeaths = xPostCovidTotal$dead - xPostCovidForecast
# Total excess deaths march-may inclusive
excessDeathsSub = excessDeaths[xPostCovidTotal$time > as.Date("2020/03/01")
  & xPostCovidTotal$time < as.Date("2020/06/01"), ]
excessDeathsInPeriod = apply(excessDeathsSub, 2, sum)
wave1_70 <- round(quantile(excessDeathsInPeriod))

# table
c <- cbind(wave1_70, wave1_total)
table3 <- c %>%
  knitr::kable(caption = "Quantiles of Posterior Sample Excess Deaths,
    March 1st - May 31st, 2020" , booktabs = T, digits = 6,
    col.names = c('70+ Years of Age', 'Total')) %>%
  kableExtra::kable_classic(full_width = F, html_font = "Cambria") %>%
  kableExtra::kable_styling(latex_options = c("HOLD_position", "striped"))
table3

### HYPOTHESIS 2
par(mfrow = c(1, 2))

## 70+
# Take posterior samples of the intensity
sampleList = INLA::inla.posterior.sample(100, res70,
  selection = list(Predictor = 0))

```

```

sampleIntensity = exp(do.call(cbind,
                             Biobase::subListExtract(sampleList, "latent")))
sampleDeaths = matrix(rpois(length(sampleIntensity),
                             sampleIntensity), nrow(sampleIntensity),
                      ncol(sampleIntensity))
# calculate excess deaths
xPostCovidTotal = xPostCovid[xPostCovid$age == "70 years old and over", ]
xPostCovidTotal = xPostCovidTotal[xPostCovidTotal$time > as.Date("2020/09/01")
                                & xPostCovidTotal$time < as.Date("2020/10/21"),]
xPostCovidForecast = sampleDeaths[match(xPostCovidTotal$time,
                                         xForInlaTotal70$time), ]
excessDeaths = xPostCovidTotal$dead - xPostCovidForecast
# plot samples of excess deaths
temp70 <- round(quantile(excessDeaths))
matplot(xPostCovidTotal$time, excessDeaths, type = "l", lty = 1,
        col = "#00000030",
        main = 'Figure 4.: Posterior Sample Excess Deaths Among 70+ yo',
        xlab = 'Time', ylab = 'Excess Death', cex.main = 0.7)

## < 50
# fit a model for 70+ yo deaths in INLA
xForInlaTotal50= xForInla[xForInla$age == '0-49 years old', ]
# model
res50 = inla(dead ~ sin12 + sin6 + cos12 + cos6 +
             f(timeId, prior='pc.prec', param= c(log(1.2), 0.5)) +
             f(timeForInla, model = 'rw2', prior='pc.prec',
               param= c(0.01, 0.5)),
             data=xForInlaTotal50,
             control.predictor = list(compute=TRUE, link=1),
             control.compute = list(config=TRUE), family='poisson')
# Take posterior samples of the intensity
sampleList = INLA::inla.posterior.sample(100, res50,
                                         selection = list(Predictor = 0))
sampleIntensity = exp(do.call(cbind,
                             Biobase::subListExtract(sampleList, "latent")))
sampleDeaths = matrix(rpois(length(sampleIntensity),
                             sampleIntensity), nrow(sampleIntensity),
                      ncol(sampleIntensity))
# calculate excess deaths
xPostCovidTotal = xPostCovid[xPostCovid$age == "0-49 years old", ]
xPostCovidTotal = xPostCovidTotal[xPostCovidTotal$time > as.Date("2020/09/01") & xPostCovidTotal$time <
                                as.Date("2020/10/21"),]
xPostCovidForecast = sampleDeaths[match(xPostCovidTotal$time,
                                         xForInlaTotal50$time), ]
excessDeaths = xPostCovidTotal$dead - xPostCovidForecast
# plot samples of excess deaths
temp50 <- round(quantile(excessDeaths))
matplot(xPostCovidTotal$time, excessDeaths, type = "l", lty = 1,
        col = "#00000030",
        main = 'Figure 5.: Posterior Sample Excess Deaths Among 0-49 yo',
        xlab = 'Time', ylab = 'Excess Death', cex.main = 0.7)

## table combined
par(mfrow = c(1, 1))

```

```

d<- cbind(temp70, temp50)
table4 <- d %>%
  knitr::kable(caption = "Quantile of Posterior Sample Excess Deaths,
    Sep 1st - Oct 21st, 2020" , booktabs = T, digits = 6,
    col.names = c('70+ Year of Age', '0-49 Years of Age')) %>%
  kableExtra::kable_classic(full_width = F, html_font = "Cambria") %>%
  kableExtra::kable_styling(latex_options = c("HOLD_position", "striped"))
table4

```