STA442 - Assignment 2

Question 1 – CO_2 emission in relation to change global economic environment

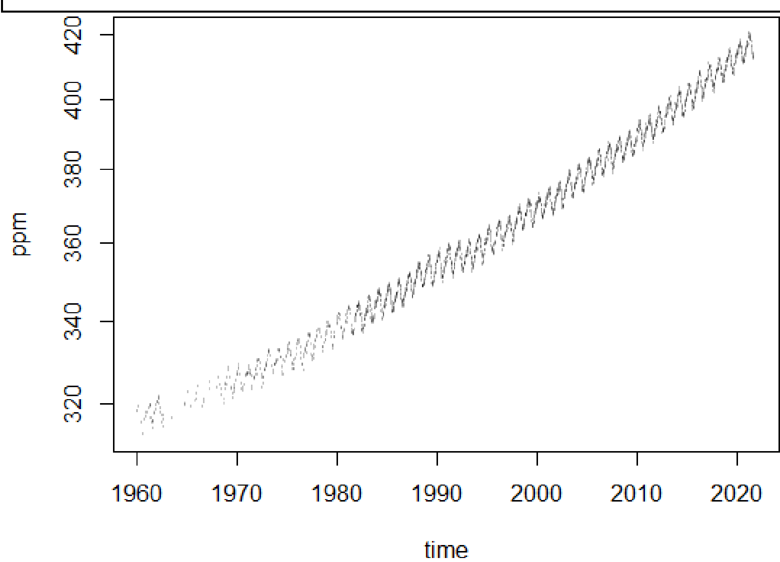
Jeff Xu
1003342545

- Part 1 – Introduction

The cause of CO_2 emissions are often considered highly relative with human activities, namely industrial productions. To further provide insight on information with bettered details on how human industrial production activities impacts on global CO_2 concentration levels, this analytic will examine through the data showing CO_2 concentrations observed from an observatory in Hawaii in contemporary times made available by the Scripps CO_2Program at *scrippsco2.ucsd.edu*. The report will demonstrate the overall changes over the contemporary period starting from 1960 that saw the beginning of industrialization from former underdeveloped states like China, to the current year 2021, and will target two specific event that is widely known to likely have notably negative impacts on the global economics and industrial productions: the fall of the Berlin wall in 1989 and the global lockdown due to COVID-19 pandemic starting in February 2020, with the hypothesis that these negative impact on the global economic will cause a slow-down in industrial production and a fall in CO_2 concentration level.

- Part 2 – model and methods

Figure 1, CO_2 concentration level observed at observatory in Hawaii



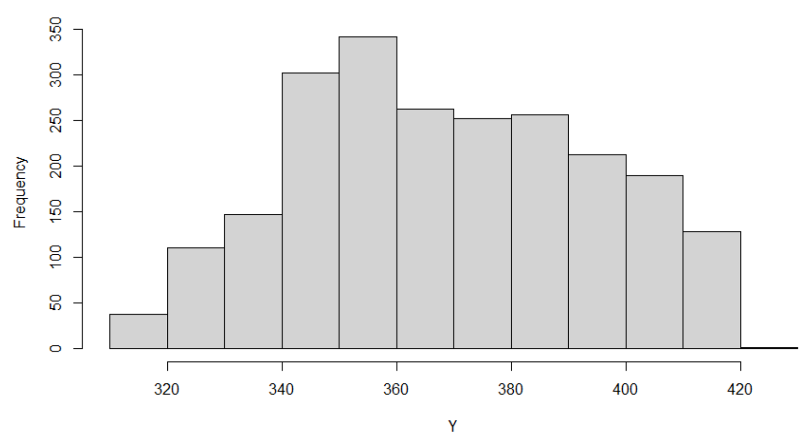


Figure 2, Histogram of frequency with Y (CO_2 concentration)

It is possible to tell from figure 1 featuring the observed CO_2 concentration level that it has a seasonal pattern, and the overall slope is going up with time. The reason of the seasonal pattern are likely due to seasonal increased production and electricity usage. Thus, the figure has a changing slope so as the random effect of this model and RW2 may be suitable to apply, and the pattern has a property of going up and back down seasonally so the property of sine and cosine can be utilized, and a generalized additive model (GAM) can be fitted as below with U_{t_i} demonstrating the random effect, X presenting the trigonometric functions for the seasonal patterns and β being the corresponding parameters for X:

$$\lambda_i = X\beta + U_{t_i}$$

And from figure 2, it the frequency of Y (CO_2 concentration) does give a patter that somehow approximately resembles with Normal distribution, which can be utilized for fitting the model with a Normal distribution model. With obvious thought, the concentration of CO_2 cannot be zero or negative even if there exists an economic decline that slows down industrial production, CO_2 emission are not likely to stop, and thus it will be continuous. We choose to use a Normal (Gaussian) distribution here as following, with Y_i as the averaged CO_2 concentration for day i.

$$Y_i \sim N(\mu, \sigma^2)$$

$$U_{t_i} \sim N(0, \sigma^2)$$

The pc.prec penalized complexity prior is applied here and the standard deviation σ^2 for both Y_i and V_j can be set with:

$$Prob(\sigma > 0.1) = 0.5$$

The prior mean would be changes in mean values in a 0.10 ratio. The 4 covariances are defined by trigonometric functions for the seasonal trending mentioned above, which are cos12, sin12, cos6, sin6.

$$cos12 = cos(2\pi \times \text{co2s\$timeYears})$$

$$sin12 = sin(2\pi \times \text{co2s\$timeYears})$$

$$cos6 = cos(2 \times 2\pi \times \text{co2s\$timeYears})$$

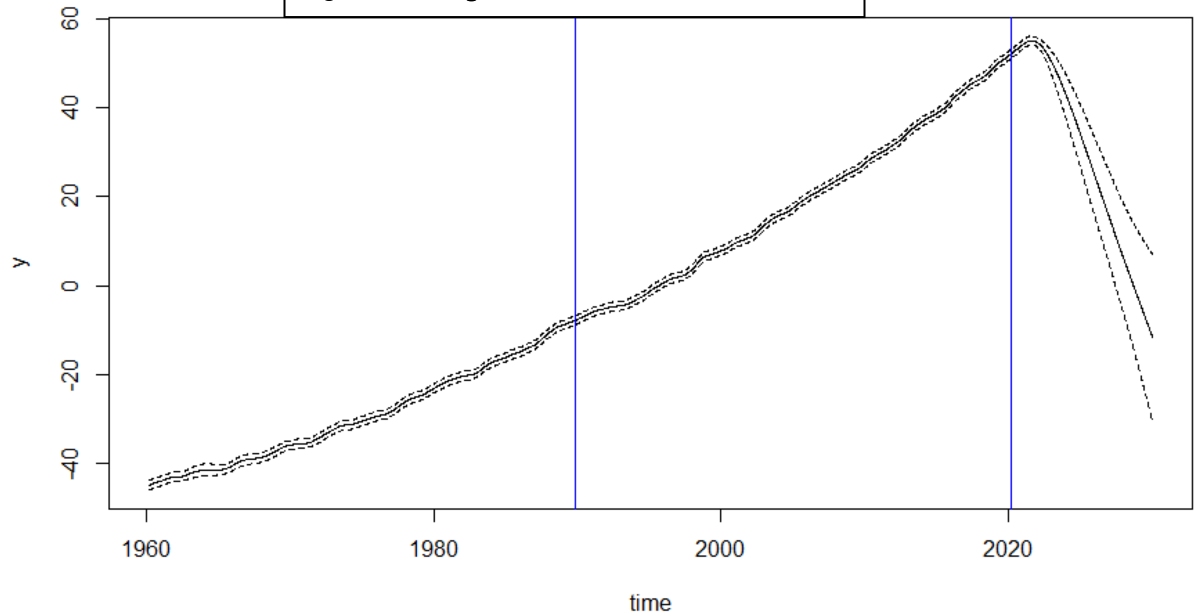
$$sin6 = sin(2 \times 2\pi \times \text{co2s\$timeYears})$$

And the distribution are as follows:

$$[U_1 \dots U_T]^T \sim RW2(0, \sigma_U^2)$$

- Part 3 – Results

Figure 3, Change of random effect with time



As shown in above, plots with the blued lines presenting the two specific timing that our hypothesis assumed to have reduced industrial activity. In figure 3 which shows the change of random effect with time which is the change of the random slope with time, we can tell the random effect generally increases with time, which makes sense since global industrialization increases with time, considering cases like China which industrializes rapidly during the time span onwards from 1960. From here we can also tell that the change of random effect does not seem to reduce or slowdown its rate of increase right after the two events, but rather a while after the events happen as after the 1989 line that the plot still went up at a usual rate, but slowed down its rate of going after a while after, and for the 2020 case, the plot is still going upwards at a rapid rate but experience a prediction of a huge drop a while after.

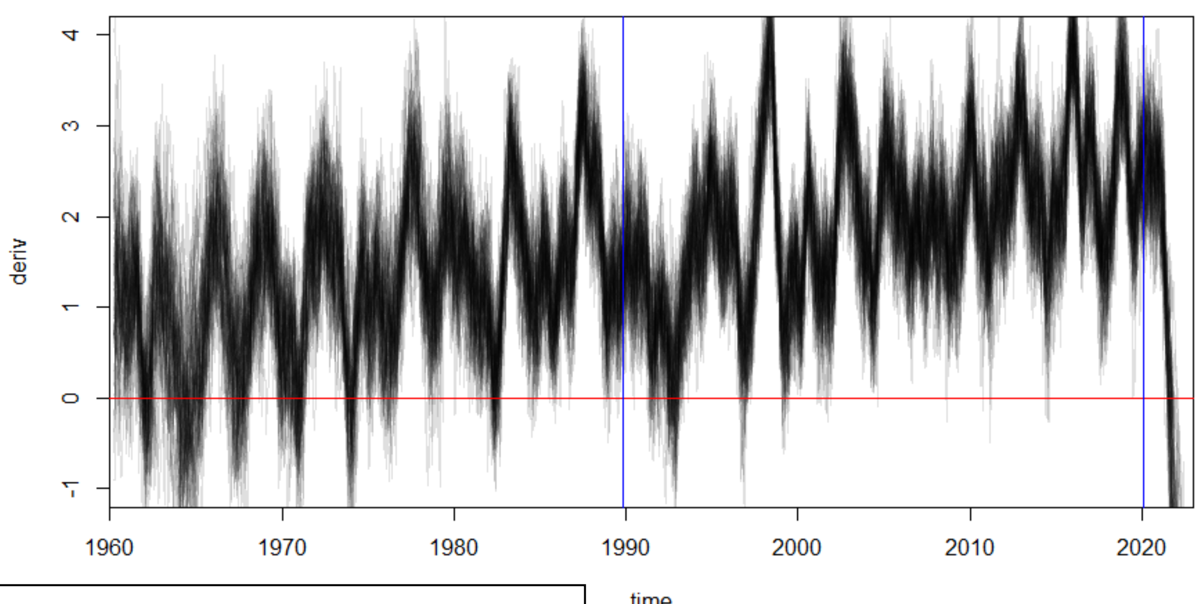


Figure 4, Derivative of CO_2 concentration with time

And in figure 4, that the derivative of CO_2 on the two occasions were decreasing despite of the seasonal pattern that it should be raising, which is in unison with what was shown earlier in the previous plot that the rate that the slope increases slows down. Which we can assume that the overall CO_2 concentration level are not likely to reduce as soon as industrial activities reduce, rather, the rate of increase of CO_2 emission will slow down, which can be explained by the decrease in derivative here implies that the emission of CO_2 has slowed down during those two occasions.

- Part 4 – Summary

Considering the case as the former underdeveloped nation in Asia, Latin America and Africa industrialize at an extremely rapid rate in recent years, these new emergent states combine with the already existing industrialized states causes further problem to the current existing difficult CO_2 concentration problem that leads to global warming. There exist opportunities for us to gain a great foresight into the future through analyzing the historical data of the CO_2 concentration level, and an insight in terms of the performance of global economy, and an hourglass that give us warning ahead of time countdowns for the consequences of global warming.

First from the observation data provided by an observatory in Hawaii that we found that the CO_2 emission has a seasonal pattern likely due to seasonal increased production and electricity usage. And by fitting the data with GAM modelling with a normal model and trigonometric function for the seasonal pattern, we can observe from the change of random effect that the overall CO_2 emission is increasing every year, likely due to the industrial expansion in developed nation and the rapid industrialization of states in the developing world, and together with the derivative graph, it shows us that the effect on CO_2 concentration level from a disastrous economic effect tends appear a little while after the event has occurred rather than instantly, and it would not reduce the overall CO_2 level, but rather slow down its rate of increasing.

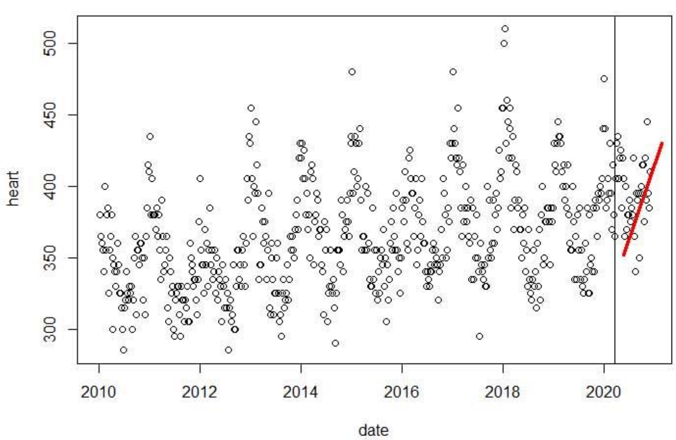
Question 2 – Cause of death of other diseases in relation with COVID-19

- Part 1 – Introduction

As the COVID-19 pandemic spreads globally in 2020, leading to introduction of nation policies on lockdowns and quarantines, which changes the social environments significantly as measures to contain the COVID disease were installed such as social-distancing, mask-wearing, and changes in global environments such as the decreased accessibility to hospitals as influx of large volumes of patients infected with COVID, and decreased air pollution due to slow down on industrial production as the global economic are impacted negatively by the pandemic.

In such a giant change of social environment, it is possible to consider the impact on human mortality rates are impacted hugely as well, with the data provided by Statistics Canada at *https://open.canada.ca/data/en/dataset/aed00edc-26ad-414c-8aa3-82212059ef8a*, this report will feature the quantification of the excess or deficit of mortality in the Canadian province Ontario consisting from four types of mortality causes of some two contagious causes: accidents and chronic lower respiratory diseases, and two non-contagious fatal diseases: Malignant neoplasms (cancers), Diseases of the heart (heart attacks), with a hypothesis that the mortality from the two contagious causes will decrease as the installation of the said pandemic measures, and an increase of mortality due to the two non-contagious fatal causes as the accessibility to healthcare are limited during the COVID pandemic.

- Part 2 – model and methods



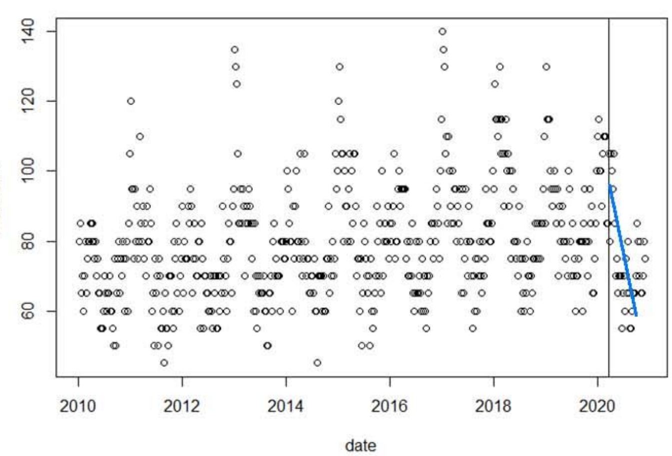
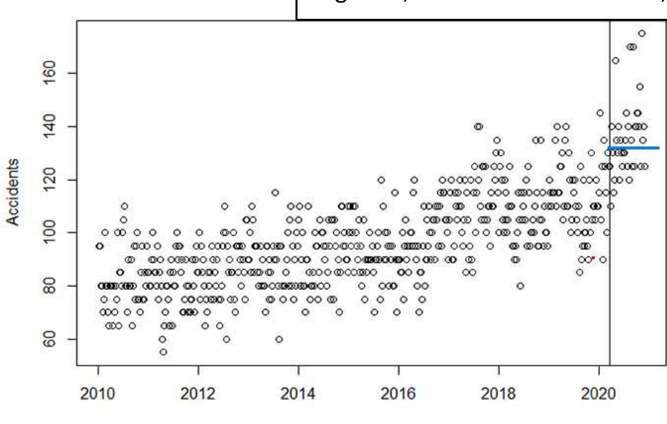
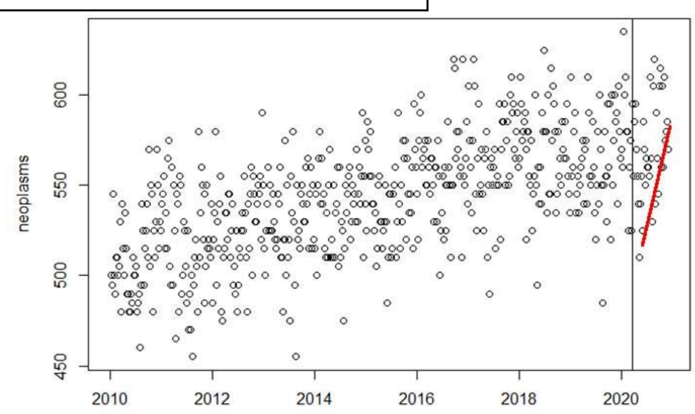


Figure 1, Trend of Heart disease, Respiratory, Accidents and Neoplasms





As this is an analysis of COVID-19's effect on the mortality rate in Ontario, through common logic, the data set would be positive and discrete, and it can be considered that the people losing their life along time are likely to follow a Poisson distribution, and the number of deaths in a period is Y_i will be assumed to have distribution of following:

$$Y_i \sim Poisson(\lambda_i)$$

And as shown above in figure 1, trend for all 4 types of mortality causes, there were also seasonal patterns of mortality rate going up and down existing in them. Thus, a semi-parametric model can be fitted, as shown below:

$$\log(\lambda_i) = X_i\beta + U_{t_i} + V_i$$

Here λ_i represent the death count expect in a period i in Ontario, and it is presented in it's logged mortality rate with $\log(\lambda_i)$. The trigonometric functions defining the seasonal pattern would be presented by X_i, and β would be the β being the corresponding parameters for X.

The four kinds of covariates defined by sine and cosine trigonometric functions referred to by X_i including sin12, cos12, sin6, and cos6, are shown below:

$$cos12 = cos(2\pi \times \text{x\$dateInt} \div 365.25)$$

$$sin12 = sin(2\pi \times \text{x\$dateInt} \div 365.25)$$

$$cos6 = cos(2 \times 2\pi \times \text{x\$dateInt} \div 365.25)$$

$$sin6 = sin(2 \times 2\pi \times \text{x\$dateInt} \div 365.25)$$

X$dateInt is the variable presenting the integer form of time in days , and the time variable is set with Jan 1, 2015, as its origin, and 365.25 days as one year. The parametric portion U_i would be following a Gaussian distribution and it gives:

$$U_{t_i} \sim N(0, \sigma_U^2)\ \sigma_V^2$$

For the non-parametric random effect V_i will have follow RW2 as the following distribution in general:

$$[V_1 \ldots V_T]^T \sim RW2(0, \sigma_V^2)$$

The pc.prec penalized complexity prior is applied here and the standard deviation σ^2 for both Y_i and V_j can be set with:

$$Prob(\sigma_V > 100) = 100^{-2}$$

- Part 3 – Results

From figure 1 showing the raw data, it is possible to tell there exists a seasonal pattern, and from the line showing the time of March 2020, number of deaths for heart disease and neoplasms are rising depicted by red line, and decreasing for respiratory disease depicted by blue line, and a no specific trend of rising nor decreasing for accident deaths also depicted by blue line.

-Heart disease

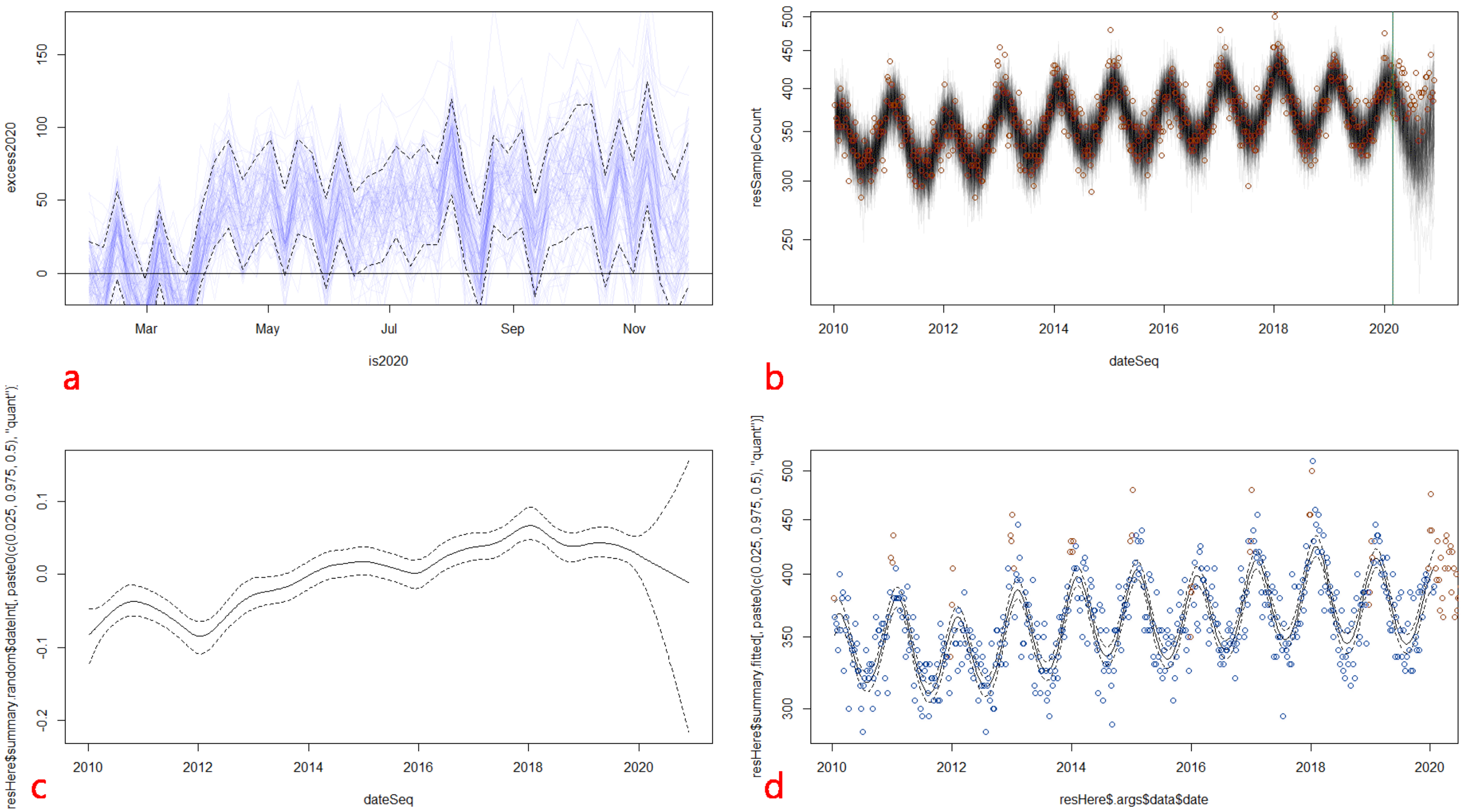


Figure 2, info of the fitted model of heart diseases, a) excess death and confidence interval, b) sample count along time, c) random effect along time, d) number of deaths along time

As shown in above figure, figure 2a) shows the excess death and its confidence interval. The confidence interval had its lower bound passing the solid black line presenting 0, and there is a constant excess of death after mid-March 2020, which tells us that there are more people are losing their life than we have estimated, implies that the effect on heart disease deaths due to society effect of COVID are a lot more severe than we have anticipated.

Quantile Table for Excess Deaths for Heart Diseases					
Quantile	0%	25%	50%	75%	100%
Excess Death	-3302.00	1844.00	3639.50	5383.25	8022.00

Above quantile table it is for excessive deaths as to gain some bettered information than from just the graph, and as a sidenote, its possible for the numbers in that chart to go below 0, as it just means that we overestimated the number of deaths. We would wish for in a good scenario, that the median, which is the 50% quantile to stay close to and stable at around 0 so it is accurate, with as small a distance between the 25% and 75% quantile, so it is precise, not the 0% and 100% since there are cases of outliers. Coming back to the table, the difference between the 25% quantile and 75% quantile is okay, about 3539.25, and the 50% quantile is at 3639 which quite high, implies that there is a lot of excessive death that we did not anticipate that our model tends to underestimate the number of deaths, and an okay distance between 25% and 75% quantile implies that our estimations is precise at an okay level.

Also, on figure 2c) it shows the change of random effect with time, there isn’t really a particular pattern, but over the time spam, it tends to be increasing with time, and it is on a trend of decrease somewhere before 2020, but as shown earlier from the excess deaths, that we underestimated the deaths after 2020. And from figure 2b) and 2d), it also shows that there is a seasonal pattern, and the death counts are somehow increasing with time.

-Neoplasms

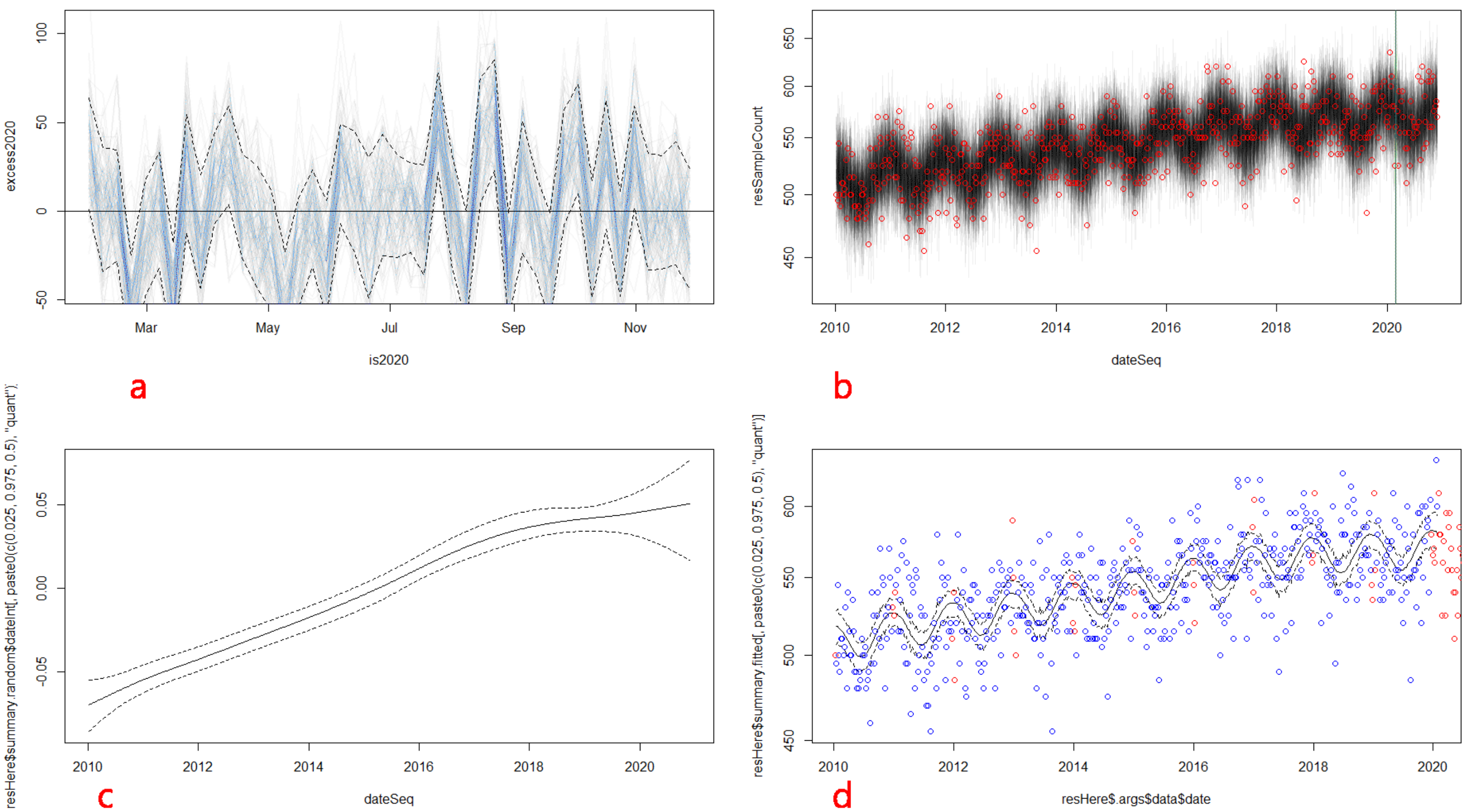


Figure 3, info of the fitted model of neoplasms, a) excess death and confidence interval along time, b) sample count along time, c) random effect with time, d) number of deaths along time

In 3a) of above figure shows the excess death and its confidence interval, overall the model can said to be stable around 0, there are a few occasions before March 2020 and around May 2020 that the upper bound went below zero which we have overestimated and there are lesser death than we have anticipated, and starting from July 2020, more excess of death which above 0 begin to appear for cases with neoplasms, as the lower bound begun to spike above 0, implying that there was more death than our model expected.

Quantile Table for Excess Deaths for Neoplasms					
Quantile	0%	25%	50%	75%	100%
Excess Death	-5251.00	-1803.50	-363.00	1847.25	5225.00

The difference between the 25% quantile and 75% quantile is fine, about 3650.75, and the median is at -363, fairly close to 0 at this case, which implies that our model is fairly accurate and precise.

And on figure 3c) it shows the change of random effects along time, it is clear that random effects has been increasing since 2010 but tends to slow down the rate of increasing around 2018, likely due to the scientific improvement on the cure for neoplasms and is somehow on the raise again a little before 2020, which one can maybe argue that COVID 19 actually started by the end 2019. In figure 3b) and 3c), it seems like that there is a seasonal pattern, and both indicates the deaths from neoplasms has been increasing with time.

-Respiratory Diseases

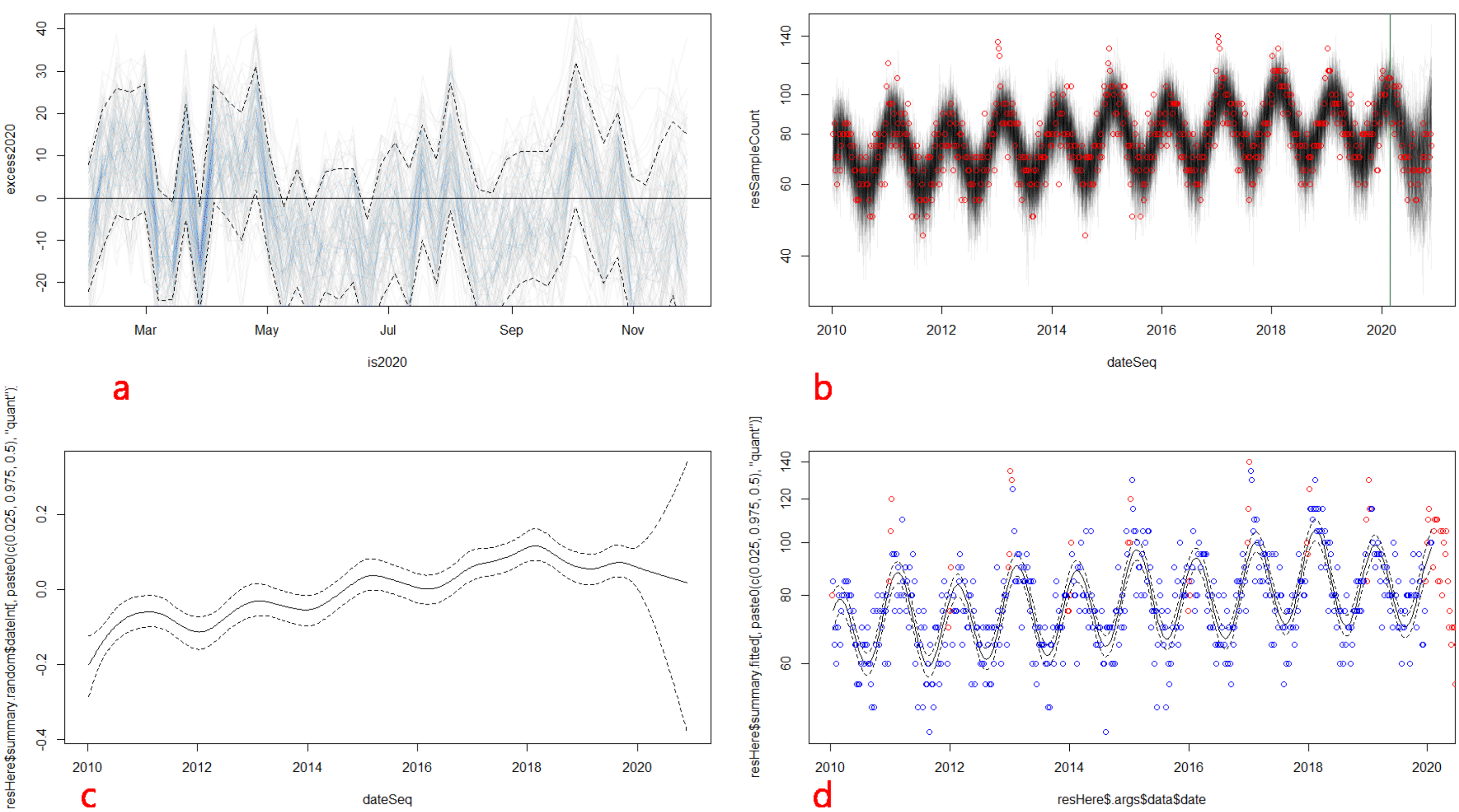


Figure 4, info of the fitted model of respiratory diseases, a) excess death and confidence interval along time, b) sample count along time, c) random effect along time, d) number of deaths along time in figure 4a) it can be somehow noted that its upper bound has been lower than zero after arrival of March 2020, which implies we have overestimated the deaths by a small bit and the situation is a little bit lesser severe, but overall it is stable around 0, so our estimation is quite accurate.

Quantile Table for Excess Deaths for Neoplasms					
Quantile	0%	25%	50%	75%	100%
Excess Death	-2156.00	-1342.25	-709.00	296.00	1446.00

The difference between the 25% quantile and 75% quantile is close about 1638.25 so our model is very precise, and the median is at -709, a okay close to 0, so that our model is fairly accurate and very precise.

In figure 4c) shows the change of random effect along with time, which we can tell that it has been increasing since 2010 but begun to reduce defying its increasing pattern somewhere a little before 2020, we can again argue that COVID actually begun by the end of 2019, and people who are careful would have already taken counter measures such as putting on face masks. And from 4b) and 4d) we see that there is a seasonal pattern on the death counts, and it seems like it's over all slope is somewhat increasing with time.

-Accident Deaths

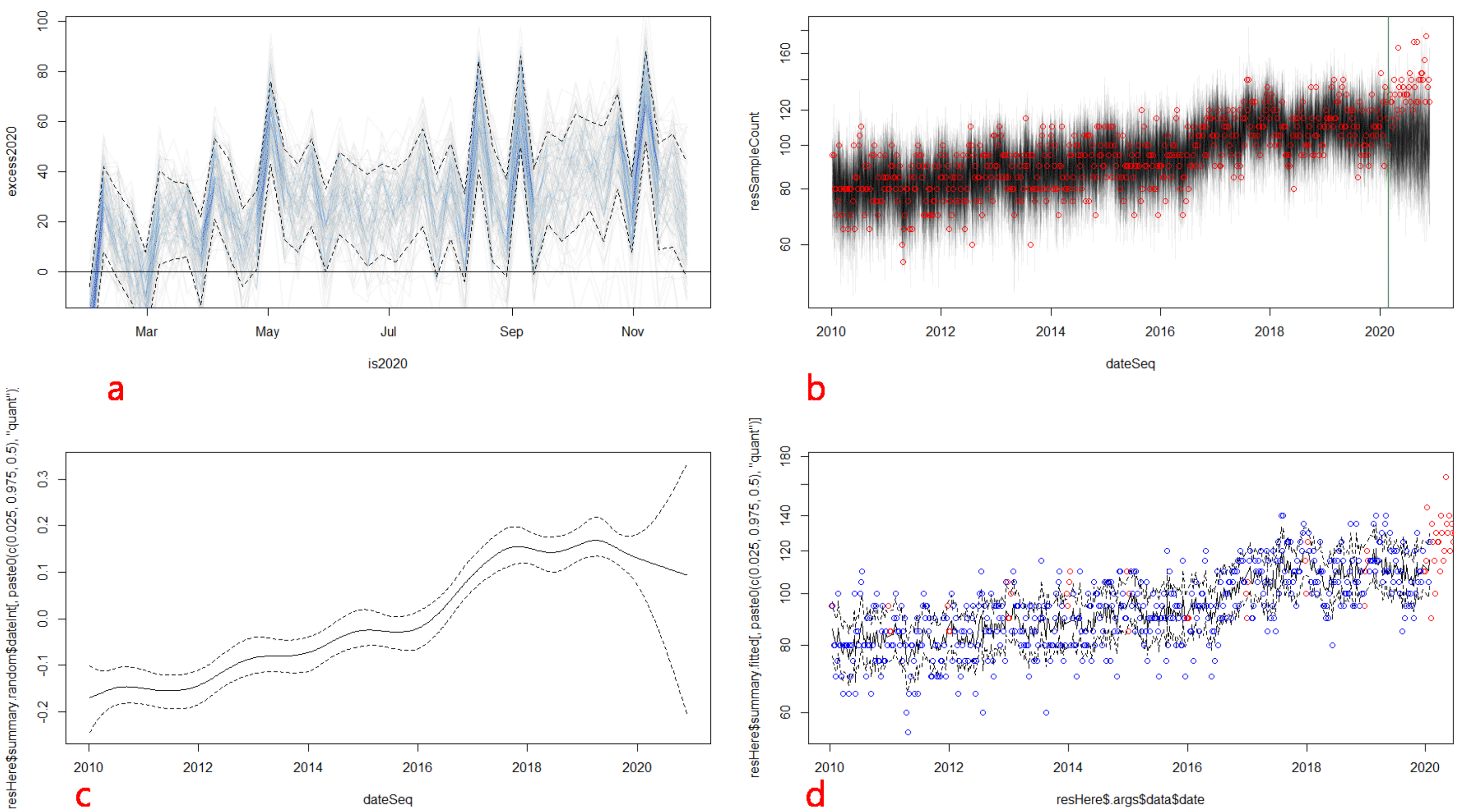


Figure 5, info of the fitted model of accidents, a) excess death and confidence interval along time, b) sample count along time, c) random effect along time, d) number of deaths along time

In figure 5a) shows the excess death and its confidence interval, it is clear that after the arrival of March 2020, that the lower bound has rapidly been above 0, which implies that after March, we underestimated the deaths by accidents substantially and the situation was more as severe.

Quantile Table for Excess Deaths for Neoplasms					
Quantile	0%	25%	50%	75%	100%
Excess Death	-2355.0	1768.5	2649.5	3441.0	6914.0

The difference between the 25% quantile and 75% quantile is very close, about 1672.5 that our model is very precise, and the median is quite high at 2649.5, so our model very precise but tends to underestimate the amount of deaths caused by accidents.

In figure 5c) shows the change of random effect along with time, which shows that the death by accidents has been increasing with time, over the longer period from 2010 it spiked high after 2016 and begun to decrease somewhere by the end of 2019. From figure 5b) and 5d), it shows that the death counts has been increasing with time.

Part 4 – Summary

With consideration of the COVID-19 pandemic, we made hypothesis on that the mortality caused by two contagious causes: accidents and chronic lower respiratory diseases will decrease, and mortality rate of two non-contagious fatal diseases: Malignant neoplasms (cancers), Diseases of the heart (heart attacks) will increase. From the analysis of the raw data, it showed that there is a seasonal pattern for all 4 causes, and from the timeline marking March 2020, heart disease and neoplasms has a trend of rising, respiratory disease with a trend of decreasing, and accident deaths has a trend of no specific change.

The fitted model showed to us that there is an excess of death for heart disease, neoplasms and accidents starting somewhat around the time interval that COVID was spread, which implies that there are much greater number of deaths from them after COVID than we have estimated. And we find respiratory disease model to be over all stable around 0, and some small negative excessive death, which implies that our model is fairly accurate with some small overestimations.

The random effect of the model is increasing over time for all 4 causes, and for heart disease it tends to decrease sometime before 2020 but shown earlier we underestimated the deaths after 2020 so it is likely more severe, and for neoplasms it tends to have slowed down rate of increasing after 2018 likely due to bettered technology for curing the disease, then with a trend of going up again after 2020. The random effect for respiratory diseases tends to decrease after 2020 against its trend of going high with time, and for accident deaths, over the longer period from 2010 it spiked high after 2016 and tends to be on the decrease after somewhere from 2019.

From the death counts death counts of the data or fitted model of all 4 mortality causes are all increasing with time, however, consider the population of Ontario is also increasing with time, it make sense the overall death counts increases. Neoplasms model was stable despite some excessive deaths, it had a raw data trend of increasing, information from random effects shows that it is increasing, and for heart disease despite its random effect was decreasing, but we from excess death information we underestimated it and is more severe, so one can think that the limit on healthcare caused COVID have an effect on increasing the mortality rate of heart or neoplasms diseases, and our hypothesis tends to be correct about these two causes of death. As the raw data trend, excessive deaths and random effects decreased for respiratory diseases all decreased, the COVID policies regarding to mask wearing, social distancing, and better hygiene practices was effective helping to reduce mortality rate of respiratory diseases too, which lies in unison with our hypothesis. And lastly for the accident deaths, we have underestimated the number of deaths, the random effect is on a trend of decreasing, so it can be said that the deaths caused by accidents is decreasing, but not as severe as we anticipate it to be, which still goes in agreement with our hypothesis. In conclusion, our hypothesis was proven to be correct, that as the arrival of COVID, the limit healthcare availability did cause an increase of mortality on fatal non-contagious diseases, and the counter pandemic measures such as bettered hygiene practices, social distancing and mask wearing implanted by the government tends to lower the spread of respiratory diseases, and together with avoiding going outdoor as a public response against COVID also reduces the deaths caused by accidents.

Appendix

```
################################################################################
################################## QN1 ######################################
################################################################################
#CO2
cUrl = paste0("http://scrippsco2.ucsd.edu/assets/data/atmospheric/",
          "stations/flask_co2/daily/daily_flask_co2_mlo.csv")
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$date = as.Date(co2s$day)
co2s$time = strptime(paste(co2s$day, co2s$time), format = "%Y-%m-%d %H:%M",
```

```
    tz = "UTC")

# remove low-quality measurements
co2s = co2s[co2s$quality == 0, ]
plot(co2s$date, co2s$co2, log = "y", cex = 0.3, col = "#00000040",
    #main = "Figure 1",
    xlab = "time", ylab = "ppm")
plot(co2s[co2s$date > as.Date("2015/3/1"), c("date",
    "co2")], log = "y", type = "o", xlab = "time",
    ylab = "ppm", cex = 0.5)

hist(co2s$co2, xlab = "Y", ylab = "Frequency", main = '')

#The code below might prove useful.
co2s$dateWeek = as.Date(lubridate::floor_date(co2s$date,unit = "week"))
co2s$timeYears = as.numeric(co2s$date)/365.25
co2s$cos12 = cos(2 * pi * co2s$timeYears)
co2s$sin12 = sin(2 * pi * co2s$timeYears)
co2s$cos6 = cos(2 * 2 * pi * co2s$timeYears)
co2s$sin6 = sin(2 * 2 * pi * co2s$timeYears)

allDays = seq(from = min(co2s$dateWeek), to = as.Date("2030/1/1"),
    by = "7 days")
table(co2s$dateWeek %in% allDays)

co2s$dateWeekInt = as.integer(co2s$dateWeek)
library("INLA", verbose = FALSE)

# disable some error checking in INLA
mm = get("inla.models", INLA:::inla.get.inlaEnv())
if (class(mm) == "function") mm = mm()
mm$latent$rw2$min.diff = NULL
assign("inla.models", mm, INLA:::inla.get.inlaEnv())
co2res = inla(co2 ~ sin12 + cos12 + sin6 + cos6 + f(dateWeekInt,
    model = "rw2", values = as.integer(allDays), prior = "pc.prec",
    param = c(0.001, 0.5), scale.model = FALSE), data = co2s,
    family = "gaussian", control.family = list(hyper = list(prec = list(prior = "pc.prec",
    param = c(1, 0.5)))), control.inla = list(strategy = "gaussian"),
    control.compute = list(config = TRUE), verbose = TRUE)
qCols = c("0.5quant", "0.025quant", "0.975quant")
1/sqrt(co2res$summary.hyperpar[, qCols])

#0.5quant 0.025quant 0.975quant
#Precision for the Gaussian observations 0.6078543957 0.6251706035 0.5910606757
#Precision for dateWeekInt 0.0001653206 0.0001937264 0.0001428522
matplot(co2res$summary.random$dateWeekInt[, qCols],
    type = "l", lty = 1)
# source('https://bioconductor.org/biocLite.R')
# biocLite('Biobase')
sampleList = INLA::inla.posterior.sample(50, co2res)
sampleMat = do.call(cbind, Biobase::subListExtract(sampleList,
    "latent"))
sampleMean = sampleMat[grep("dateWeekInt", rownames(sampleMat)),]
```

```
sampleDeriv = apply(sampleMean, 2, diff) * (365.25/7)
forSinCos = 2 * pi * as.numeric(allDays)/365.25
forForecast = cbind(`(Intercept)` = 1, sin12 = sin(forSinCos),
    cos12 = cos(forSinCos), sin6 = sin(2 * forSinCos),
    cos6 = cos(2 * forSinCos))
forecastFixed = forForecast %*% sampleMat[paste0(colnames(forForecast),
    ":1"), ]
forecast = forecastFixed + sampleMean
matplot(allDays, forecast, type = "l", col = "#00000010",
    lty = 1, log = "y", xlab = "time", ylab = "ppm")
forX = as.Date(c("2018/1/1", "2025/1/1"))
forX = seq(forX[1], forX[2], by = "1 year")
toPlot = which(allDays > min(forX) & allDays < max(forX))
matplot(allDays, forecast, type = "l", col = "#00000020",
    lty = 1, log = "y", xlab = "time", ylab = "ppm",
    xaxs = "i", xaxt = "n", xlim = range(forX),
    ylim = range(forecast[which.min(abs(allDays - max(forX))), ]))

points(co2s$date, co2s$co2, col = "red", cex = 0.3)
axis(1, as.numeric(forX), format(forX, "%Y"))

matplot(allDays, co2res$summary.random$dateWeekInt[,
    qCols], type = "l", col = "black", lty = c(1, 2,
    2), xlab = "time", ylab = "y")
abline(v = as.numeric(as.Date("1989-11-15")) , col = "blue")
abline(v = as.numeric(as.Date("2020-02-15")) , col = "blue")

matplot(allDays[-1], sampleDeriv, type = "l", lty = 1,
    xaxs = "i", col = "#00000020", xlab = "time", ylab = "deriv",
    xlim = as.Date(c("1960/1/1", "2023/1/1")),
    ylim = c(-1, 4))
    #ylim = quantile(sampleDeriv, c(0.025, 0.995)))
abline(h = 0 , col = "red")
abline(v = as.numeric(as.Date("1989-11-15")) , col = "blue")
abline(v = as.numeric(as.Date("2020-02-15")) , col = "blue")

matplot(allDays[toPlot], sampleDeriv[toPlot, ], type = "l",
    lty = 1, lwd = 2, xaxs = "i", col = "#00000020",
    xlab = "time", ylab = "deriv", xaxt = "n", ylim = quantile(sampleDeriv[toPlot,
    ], c(0.01, 0.995)))
axis(1, as.numeric(forX), format(forX, "%Y"))


###############################################################################
###################################  QN2  ######################################
###############################################################################

deadFile = Pmisc::downloadIfOld("https://www150.statcan.gc.ca/n1/tbl/csv/13100810-eng.zip",)
                    #path = "../data")
(deadFileCsv = deadFile[which.max(file.info(deadFile)$size)])
x = read.csv(deadFileCsv)
x[1:2, ]
```

```
x$date = as.Date(as.character(x[[grep("DATE", names(x))]]))
x$province = gsub("[,].*", "", x$GEO)
# remove 2021 data, which appears incomplete
x = x[x$date < as.Date("2020/12/01") & x$province ==
    "Ontario", ]

###hist(x$dateInt, xlab = "Y", ylab = "Frequency", main = '')

#for (D in c("heart", "neoplasms", "respiratory", "Accidents")) {
#  plot(x[grep(D, x$Cause), c("date", "VALUE")], ylab = D)
#  abline(v = as.Date("2020/03/17"))
#}

dateSeq = sort(unique(x$date))
table(diff(dateSeq))

dateSeqInt = as.integer(dateSeq)
x$dateInt = x$dateIid = as.integer(x$date)
x$cos12 = cos(2 * pi * x$dateInt/365.25)
x$sin12 = sin(2 * pi * x$dateInt/365.25)
x$sin6 = sin(2 * 2 * pi * x$dateInt/365.25)
x$cos6 = cos(2 * 2 * pi * x$dateInt/365.25)
x$dayOfYear = as.Date(gsub("^[[:digit:]]+", "0000",
                 x$date))
x$christmasBreak = (x$dayOfYear >= as.Date("0000/12/21")) |
  (x$dayOfYear <= as.Date("0000/01/12"))
xSub = x[grepl("Accidents", x$Cause, ignore.case = TRUE) &
      x$province == "Ontario", ]
xPreCovid = xSub[xSub$date < as.Date("2020/02/01") &
          (!xSub$christmasBreak), ]

library("INLA")
resHere = inla(VALUE ~ cos12 + cos6 + sin12 + sin6 +
         f(dateInt, model = "rw2", values = dateSeqInt,
          prior = "pc.prec", param = c(0.1, 0.5)) +
         f(dateIid, values = dateSeqInt, prior = "normal",
          param = c(0.1, 0.5)), data = xPreCovid,
        family = "poisson", control.compute = list(config = TRUE),
        control.predictor = list(compute = TRUE))
matplot(resHere$.args$data$date, resHere$summary.fitted[,
                     paste0(c(0.025, 0.975, 0.5), "quant")], type = "l",
    lty = c(2, 2, 1), col = "black", log = "y", ylim = range(xSub$VALUE))
points(xSub$date, xSub$VALUE, col = "red")
points(xPreCovid$date, xPreCovid$VALUE, col = "blue")
matplot(dateSeq, resHere$summary.random$dateInt[, paste0(c(0.025,
                      0.975, 0.5), "quant")], type = "l", lty = c(2,
                                   2, 1), col = "black")
toPredict = cbind(`(Intercept):1` = 1, `cos12:1` = cos(2 *
                     pi * dateSeqInt/365.25), `sin12:1` = sin(2 * pi *
                                   dateSeqInt/365.25), `cos6:1` = cos(2 * pi * dateSeqInt *
                                             2/365.25), `sin6:1` = sin(2 * pi *
dateSeqInt *
                                                    2/365.25))
```

```
dateIntSeq = paste0("dateInt:", 1:length(dateSeqInt))
dateIidSeq = paste0("dateIid:", 1:length(dateSeqInt))
resSample = inla.posterior.sample(n = 100, resHere)
resSampleFitted = lapply(resSample, function(xx) {
  toPredict %*% xx$latent[colnames(toPredict), ] +
    xx$latent[dateIntSeq, ] + xx$latent[dateIidSeq,
    ]
})
resSampleFitted = do.call(cbind, resSampleFitted)
resSampleLambda = exp(resSampleFitted)
resSampleCount = matrix(rpois(length(resSampleLambda),
                         resSampleLambda), nrow(resSampleLambda), ncol(resSampleLambda))
matplot(dateSeq, resSampleCount, col = "#00000010",
        type = "l", lty = 1, log = "y")
points(xSub[, c("date", "VALUE")], col = "red")
abline(v = as.Date("2020/03/01"), col = "green")
is2020 = dateSeq[dateSeq >= as.Date("2020/2/1")]
sample2020 = resSampleCount[match(is2020, dateSeq),
]
count2020 = xSub[match(is2020, xSub$date), "VALUE"]
excess2020 = count2020 - sample2020
matplot(is2020, excess2020, type = "l", lty = 1, col = "#0000FF10",
        ylim = range(-10, quantile(excess2020, c(0.1, 0.999))))
matlines(is2020, t(apply(excess2020, 1, quantile, prob = c(0.1,
                                                0.9))), col = "black", lty = 2)
abline(h = 0)
quantile(apply(excess2020, 1, sum))
```