The Pandemic is Over

Phil Shea

2022-07-22

Total Deaths

Total weekly death has reached back to 2019 levels, and therefore the pandemic is over. The evidence and all the calculations are below. Total death data downloaded from CDC website: https://data.cdc.gov/NCHS/Weekly-counts-of-deaths-by-jurisdiction-and-age-gr/y5bj-9g5w.

The code below downloads the latest data, and displays the structure of the data.

```
Death <- read.socrata('https://data.cdc.gov/NCHS/Weekly-counts-of-deaths-by-jurisdiction-and-age-gr/y5b

str( Death)
```

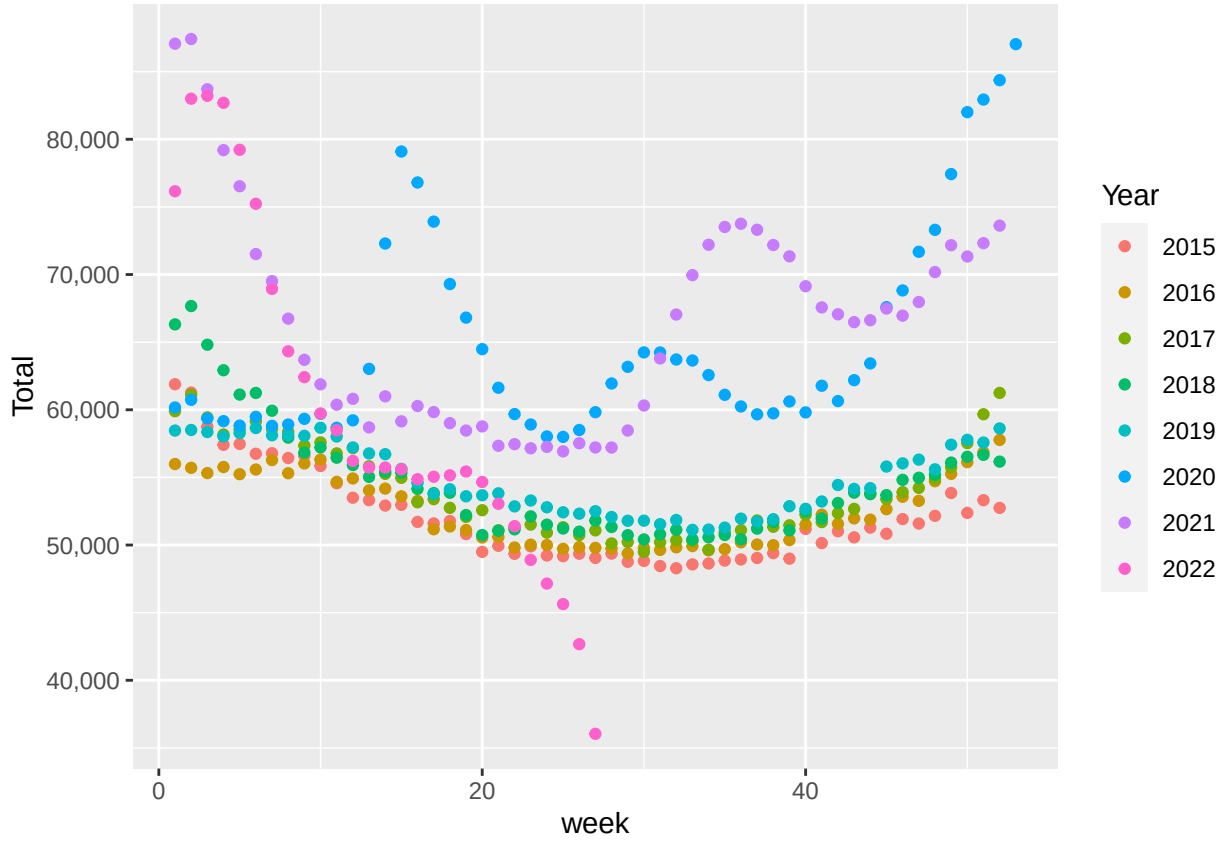
```
## 'data.frame':    231594 obs. of  11 variables:
##  $ jurisdiction      : Factor w/ 54 levels "Alabama","Alaska",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ week_ending_date  : POSIXct, format: "2015-01-10" "2015-01-17" ...
##  $ state_abbreviation: Factor w/ 54 levels "AK","AL","AR",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ year              : int  2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
##  $ week              : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ age_group         : Factor w/ 6 levels "25-44 years",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ number_of_deaths  : int  67 49 55 59 47 59 41 47 60 57 ...
##  $ time_period       : Factor w/ 4 levels "2015-2019","2020",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ type              : Factor w/ 2 levels "Predicted (weighted)",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ suppress          : Factor w/ 2 levels "","Suppressed (counts highly incomplete, <50% of expected
##  $ note              : Factor w/ 4 levels "","Data in recent weeks are incomplete. Only 60% of death
```

The data contains deaths by age group, by jurisdiction (including each state, the entire US, and New York city), and also has observations "Predicted (weighted)" (whatever that is) and "Unweighted," (tagged in the `type` column) so many deaths are double counted. The code below will filter out just the "US," "Unweighted" data, and sum the age groups to get total deaths for each week. Note also that 2020 has 53 weeks. Week 1 of 2020 started on 30 December *2019*, and week 53 covers 28 December 2020 through 3 January *2021* (this follows the rule that weeks are 7 days, and get assigned to the year with four of the seven days).

```
DS <- Death |> filter( state_abbreviation == 'US', type == 'Unweighted') |>
   group_by( week,  Year=year, date=week_ending_date) |>
   summarise( Total=sum( number_of_deaths), n=n()) # DS has weekly deaths
```

Here is the standard plot that many media outlets have been showing. It does appear that there are a remarkable number of excess deaths. The last five weeks or so typically do not have complete data yet (according to the CDC); they should be discounted.

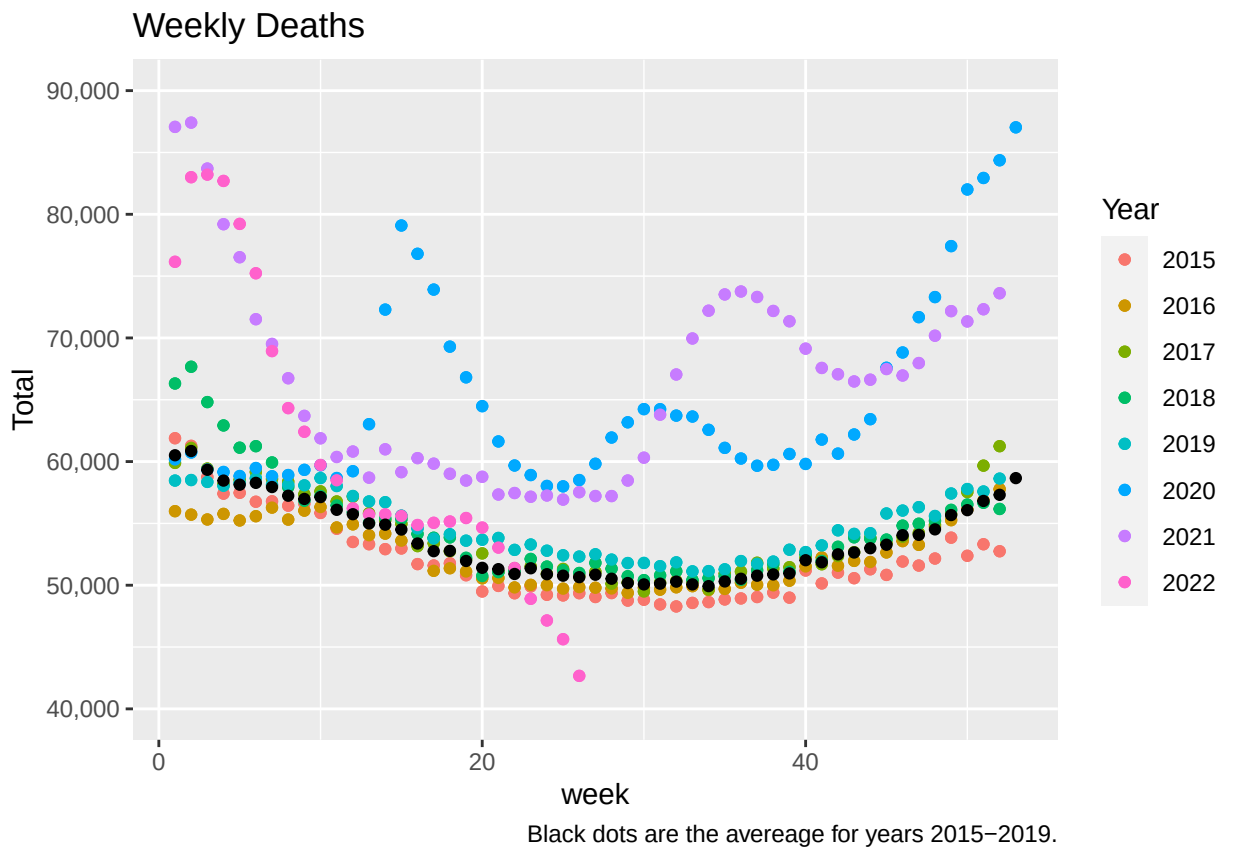
```
(dp <- ggplot( DS, mapping=aes( x=week, y=Total, color=as.factor(Year))) +
   scale_y_continuous( labels=comma) + geom_point()  + labs( color='Year'))
```



Average Deaths by Week

It is useful to see averages for non-COVID years. The code below will perform the filtering and grouping as we did earlier, but will exclude years after 2020 from the data. Now we have to deal with 2020's week 53 (it had an extra week). For simplicity's sake, we will simply make the week 53 average the mean of week 1 and week 52 average deaths.

```
DA <- Death |> # DA is average deaths by week estimate by years before 2020
   filter( state_abbreviation == 'US', type == 'Unweighted', year < 2020) |>
   group_by( week, year) |>
   summarise( Total=sum( number_of_deaths)) |>
   summarise( Mean=mean( Total), n=n())
DA <- rbind( DA, c( 53, mean( DA$Mean[ c( 1,51)]), 0)) # Deal with week 53
wmin <- DA$week[ which.min( DA$Mean)] # find minimum week in the average.
dp + geom_point() + geom_point( data=DA, aes( x=week, y=Mean), color='black') +
   scale_y_continuous( label=comma, limits=c( 40000, 90000)) +
   labs( color='Year', title = "Weekly Deaths",
         caption = "Black dots are the avereage for years 2015-2019.")
```




```
wmin <- DA$week[ which.min( DA$Mean)]
```

We cut the lower part as any sample below 40,000 is incomplete. So, the simple question is about how much higher each post 2019 week is from the average.

Shape of Weekly Deaths

In order to make reliable estimates of weekly excess deaths, we need a reliable estimate of the weekly death rate, which obviously varies by week and by year. We'll look at a polynomial approximation using an equations such as

$$D_e(w, y) = \sum_{n=0}^{4} a_n (w - w_{min})^n + by + D_{ave}$$

where D_e are the expected weekly deaths in week w and year y, w_{min} is the week number of the minimum deaths (and is 34 in this data), b is the yearly increase in deaths, and D_{ave} is the average deaths over the modeled period[1]. 2017 will be taken as year zero. A 4^{th} degree polynomial is used due to some prior work which indicates it is statistically "best."

Now we need to apply this method over the whole data set. The code below creates new `aYear` and `aw` columns, performs the linear model, and displays a summary of the fit. This is a lot of technical details, but there is some interesting details there.

[1] This could be absorbed into a_0, but we would like the polynomial coefficients to only reflect the weekly curve

```
DS$aYear <- DS$Year - 2017 # Create an adjusted year so 2017 is zero.
mtw <- mean( DS$Total[ DS$Year < 2020])
DS$aTotal <- DS$Total - mtw # The deviation from the 5 year average.
DS$aw <- DS$week - wmin # adjusted week

fit2 <- lm( aTotal ~ poly( aw, degree = 4) + aYear - 1,
            data = DS[ DS$Year < 2020,])

summary( fit2)
```

```
##
## Call:
## lm(formula = aTotal ~ poly(aw, degree = 4) + aYear - 1, data = DS[DS$Year <
##     2020, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3690.7  -398.9   -30.6   426.4  6850.1
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## poly(aw, degree = 4)1 -23335.00    1170.81 -19.931  < 2e-16 ***
## poly(aw, degree = 4)2  43293.82    1170.81  36.978  < 2e-16 ***
## poly(aw, degree = 4)3   5822.02    1170.81   4.973 1.21e-06 ***
## poly(aw, degree = 4)4  -2992.37    1170.81  -2.556   0.0112 *
## aYear                    760.98      51.34  14.822  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1171 on 255 degrees of freedom
## Multiple R-squared:  0.8877, Adjusted R-squared:  0.8855
## F-statistic: 403.1 on 5 and 255 DF,  p-value: < 2.2e-16
```

Briefly, this tells us that the coefficients are all highly significant, with the 4^{th} degree polynomial being just above 11% significance[2]. The polynomial coefficients are uninteresting[3], but the `aYear` value of `761` indicates that each year the weekly death increases by that amount (yearly: 52 * 760 or about 40,000). The R-squared values tell us that the model explains about 89% of the variation, which is pretty good.

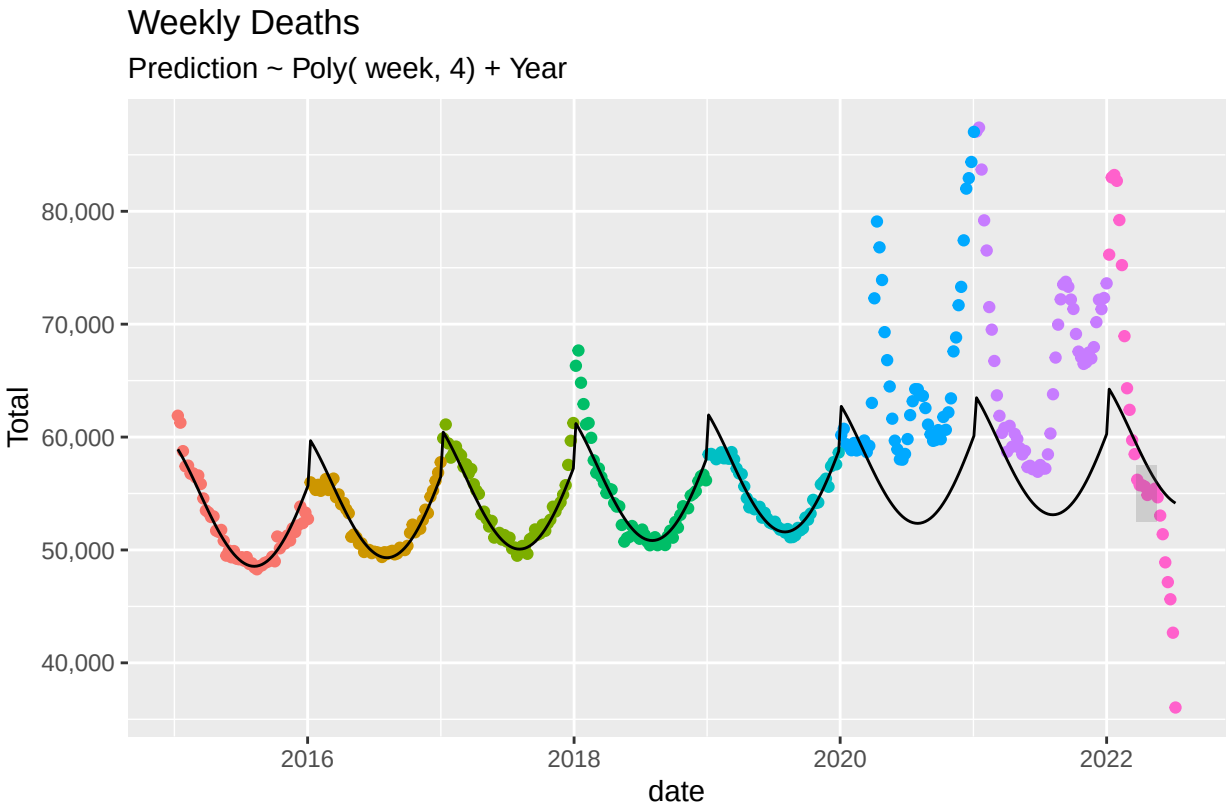
The following will extend the prediction through 2022.

```
# below will extend the prediction over the entire data frame
DS$Pred2 <- predict( fit2, newdata = DS) + mtw # must add the mean back in.
dpsw2 <- ggplot( DS, mapping = aes( x = date, y = Total)) +
   scale_y_continuous( labels = scales::comma) +
   geom_point( aes( color = factor( Year))) + labs( color = 'Year')
(dpsw3 <- dpsw2 + geom_line( data=DS,
                   mapping=aes( x  = date, y = Pred2)) +
   labs( title = "Weekly Deaths",
         subtitle = "Prediction ~ Poly( week, 4) + Year",
   caption = "Weekly deaths from CDC database.  Prediction (black line) estimated over 2015 - 2019.") +
```

[2]What this means is that there is only about an 11.2% chance that that 4^{th} degree coefficient would have happened randomly due to the data. The lower the significance the better.

[3]Principally due to the way `R` uses something called *orthogonal polynomials*

```
  annotate( "rect", xmin = as.POSIXct( "03/23/22", format = "%m/%d/%y"),
            xmax = as.POSIXct("05/17/22", format = "%m/%d/%y"), ymin = 52500,
            ymax = 57500, alpha = 0.2) + theme( legend.position = "none"))
```



Weekly deaths from CDC database. Prediction (black line) estimated over 2015 – 2019.

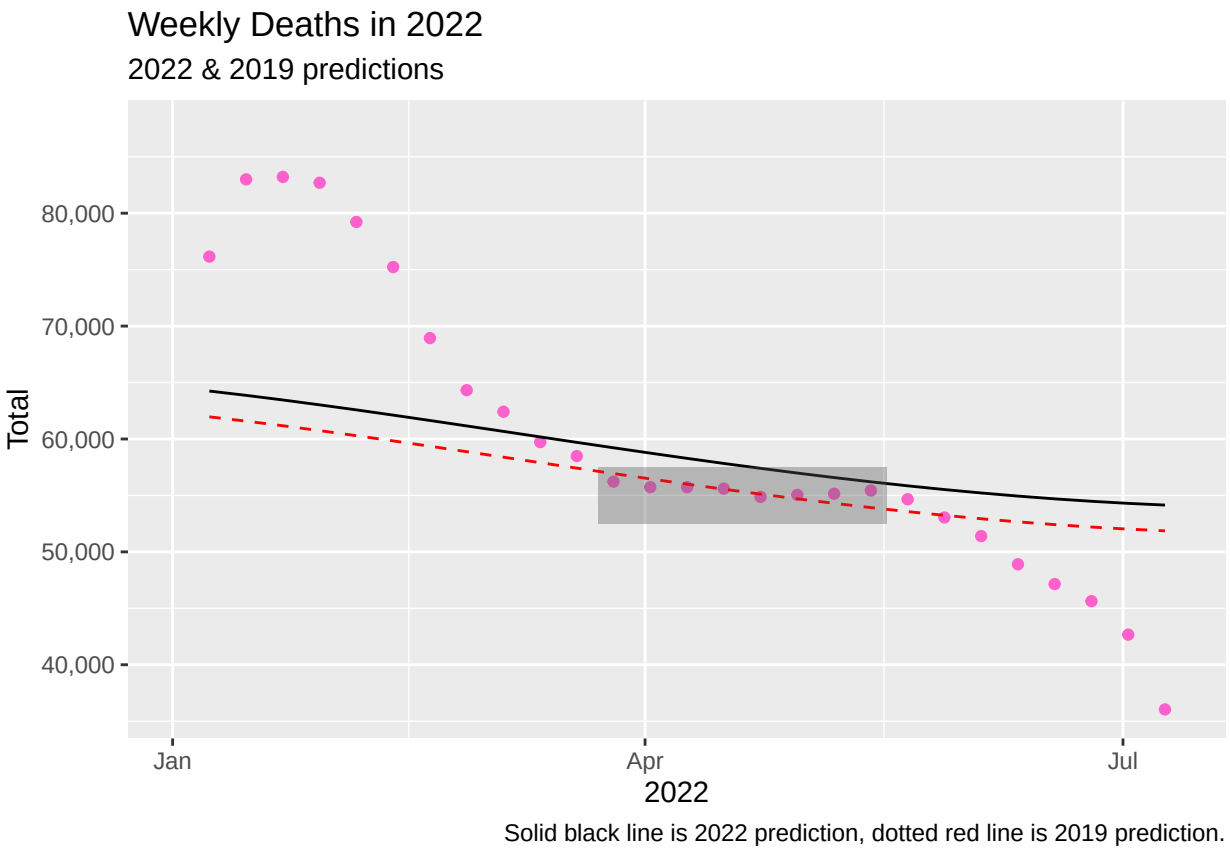
H1N1 is evident in the beginning of 2018, and drove a January peak of expected deaths. Nevertheless, the curve matches the modeled years pretty well. Note the small highlighted section in March and April of 2022.

End of the COVID 19 Pandemic

When the death curve starts to match the prediction again, the pandemic is truly over. Simply eyeballing the previous graphs, it looks like the death rate has settled back to the 2019 level. The post-pandemic expected death rate for 2022 should be lower than that predicted by the model, as many people who would have died in 2022 (had there been no pandemic) had already died. The following will demonstrate this.

```
DS2 <- DS
DS2$aYear  <- DS2$aYear - 3
DS2$Pred2 <-  predict( fit2, newdata = DS2) + mtw # must add the mean back in.
dpsw3 +  geom_line( data = DS2, mapping = aes( x = date, y = DS2$Pred2),
                    linetype=2, colour='red') +
   scale_x_datetime( limits = as.POSIXct( c("01/02/22", "07/11/22"),
                                          format = "%m/%d/%y")) +
   labs( title = "Weekly Deaths in 2022",
         subtitle = "2022 & 2019 predictions", x="2022",
 caption = "Solid black line is 2022 prediction, dotted red line is 2019 prediction.") +
   annotate( "rect", xmin = as.POSIXct( "03/23/22", format = "%m/%d/%y"),
```

```
            xmax = as.POSIXct("05/17/22", format = "%m/%d/%y"), ymin = 52500,
            ymax = 57500, alpha = 0.2) + theme( legend.position = "none")
```



Solid black line is 2022 prediction, dotted red line is 2019 prediction.

This does not mean that people are no longer dying from COVID-19. They are. Just like they have always died from the flu, from colds, and other infirmities. The point is that they are dying **just like they always have**. Like all biological systems, this may change, and there is, perhaps, evidence of a slightly increasing death rate in weeks 18 and 19 (the last two highlighted weeks). (Before you panic, look at the increase from H1N1 in 2018.) While the CDC claims that after five weeks the statistics should be reliable, experience has shown that seven weeks should be discounted, so the seventh latest point being on or below the 2019 prediction should be ignored.