Project

Aaron Banlao

Install the COVID19 Data Hub R package. Select at least two countries to work with of interest to you and make forecasts of the number of cases for the next 12 months.

Build the following models:

Build a linear regression model using TSLM(). Build an appropriate exponential smoothing model.

```
library(pacman)
p_load(fpp3, tidyverse, doParallel)
```

```
covid <- read.csv("1.csv", header = T)
```

```
head(covid)
```

```
##         id       date confirmed deaths recovered tests vaccines
## 1 8320791a 2020-03-13         2     NA        NA    NA       NA
## 2 8320791a 2020-03-14         2     NA        NA    NA       NA
## 3 8320791a 2020-03-15         2     NA        NA    NA       NA
## 4 8320791a 2020-03-16         2     NA        NA    NA       NA
## 5 8320791a 2020-03-17         8     NA        NA    NA       NA
## 6 8320791a 2020-03-18         9     NA        NA    NA       NA
##   people_vaccinated people_fully_vaccinated hosp icu vent school_closing
## 1                NA                       NA   NA  NA   NA             NA
## 2                NA                       NA   NA  NA   NA             NA
## 3                NA                       NA   NA  NA   NA             NA
## 4                NA                       NA   NA  NA   NA             NA
## 5                NA                       NA   NA  NA   NA             NA
## 6                NA                       NA   NA  NA   NA             NA
##   workplace_closing cancel_events gatherings_restrictions transport_closing
## 1                NA            NA                      NA                NA
## 2                NA            NA                      NA                NA
## 3                NA            NA                      NA                NA
## 4                NA            NA                      NA                NA
## 5                NA            NA                      NA                NA
## 6                NA            NA                      NA                NA
##   stay_home_restrictions internal_movement_restrictions
## 1                     NA                             NA
## 2                     NA                             NA
## 3                     NA                             NA
## 4                     NA                             NA
## 5                     NA                             NA
## 6                     NA                             NA
##   international_movement_restrictions information_campaigns testing_policy
## 1                                  NA                    NA             NA
```

```
## 2                                 NA                   NA             NA
## 3                                 NA                   NA             NA
## 4                                 NA                   NA             NA
## 5                                 NA                   NA             NA
## 6                                 NA                   NA             NA
##   contact_tracing facial_coverings vaccination_policy elderly_people_protection
## 1              NA               NA                 NA                        NA
## 2              NA               NA                 NA                        NA
## 3              NA               NA                 NA                        NA
## 4              NA               NA                 NA                        NA
## 5              NA               NA                 NA                        NA
## 6              NA               NA                 NA                        NA
##   government_response_index stringency_index containment_health_index
## 1                        NA               NA                       NA
## 2                        NA               NA                       NA
## 3                        NA               NA                       NA
## 4                        NA               NA                       NA
## 5                        NA               NA                       NA
## 6                        NA               NA                       NA
##   economic_support_index administrative_area_level administrative_area_level_1
## 1                     NA                         1              Grand Princess
## 2                     NA                         1              Grand Princess
## 3                     NA                         1              Grand Princess
## 4                     NA                         1              Grand Princess
## 5                     NA                         1              Grand Princess
## 6                     NA                         1              Grand Princess
##   administrative_area_level_2 administrative_area_level_3 latitude longitude
## 1                          NA                          NA       NA        NA
## 2                          NA                          NA       NA        NA
## 3                          NA                          NA       NA        NA
## 4                          NA                          NA       NA        NA
## 5                          NA                          NA       NA        NA
## 6                          NA                          NA       NA        NA
##   population iso_alpha_3 iso_alpha_2 iso_numeric iso_currency key_local
## 1       3533                                 NA                    NA
## 2       3533                                 NA                    NA
## 3       3533                                 NA                    NA
## 4       3533                                 NA                    NA
## 5       3533                                 NA                    NA
## 6       3533                                 NA                    NA
##   key_google_mobility key_apple_mobility key_jhu_csse key_nuts key_gadm
## 1                                                   NA
## 2                                                   NA
## 3                                                   NA
## 4                                                   NA
## 5                                                   NA
## 6                                                   NA
```

```
covid$id[covid$id == "59a13ceb"] <- "United States"
covid$id[covid$id == "78833522"] <- "China"
```

```
covid <- covid %>%
  filter(id == "United States" | id == "China")
head(covid)
```

```
##      id       date confirmed deaths recovered tests vaccines people_vaccinated
## 1 China 2020-01-03         0      0        NA    NA       NA                NA
## 2 China 2020-01-04         1      0        NA    NA       NA                NA
## 3 China 2020-01-05         1      0        NA    NA       NA                NA
## 4 China 2020-01-06         4      0        NA    NA       NA                NA
## 5 China 2020-01-07         4      0        NA    NA       NA                NA
## 6 China 2020-01-08         4      0        NA    NA       NA                NA
##   people_fully_vaccinated hosp icu vent school_closing workplace_closing
## 1                      NA   NA  NA   NA              0                 0
## 2                      NA   NA  NA   NA              0                 0
## 3                      NA   NA  NA   NA              0                 0
## 4                      NA   NA  NA   NA              0                 0
## 5                      NA   NA  NA   NA              0                 0
## 6                      NA   NA  NA   NA              0                 0
##   cancel_events gatherings_restrictions transport_closing
## 1             0                       0                0
## 2             0                       0                0
## 3             0                       0                0
## 4             0                       0                0
## 5             0                       0                0
## 6             0                       0                0
##   stay_home_restrictions internal_movement_restrictions
## 1                      0                              0
## 2                      0                              0
## 3                      0                              0
## 4                      0                              0
## 5                      0                              0
## 6                      0                              0
##   international_movement_restrictions information_campaigns testing_policy
## 1                                   0                    0              1
## 2                                   0                    0              1
## 3                                   0                   -1              1
## 4                                   0                   -1              1
## 5                                   0                   -1              1
## 6                                   0                   -1              1
##   contact_tracing facial_coverings vaccination_policy elderly_people_protection
## 1               1               -1                  0                         0
## 2               1               -1                  0                         0
## 3               2               -1                  0                         0
## 4               2               -1                  0                         0
## 5               2               -1                  0                         0
## 6               2               -1                  0                         0
##   government_response_index stringency_index containment_health_index
## 1                      5.99            0.00                     6.85
## 2                      5.99            0.00                     6.85
## 3                     10.68            2.78                    12.20
## 4                     10.68            2.78                    12.20
## 5                     10.68            2.78                    12.20
## 6                     10.68            2.78                    12.20
##   economic_support_index administrative_area_level administrative_area_level_1
## 1                      0                         1                       China
## 2                      0                         1                       China
## 3                      0                         1                       China
## 4                      0                         1                       China
```

```
## 5                      0                         1                      China
## 6                      0                         1                      China
##   administrative_area_level_2 administrative_area_level_3 latitude longitude
## 1                          NA                          NA 32.82838  111.6491
## 2                          NA                          NA 32.82838  111.6491
## 3                          NA                          NA 32.82838  111.6491
## 4                          NA                          NA 32.82838  111.6491
## 5                          NA                          NA 32.82838  111.6491
## 6                          NA                          NA 32.82838  111.6491
##   population iso_alpha_3 iso_alpha_2 iso_numeric iso_currency key_local
## 1 1392730000         CHN          CN         156          CNY        NA
## 2 1392730000         CHN          CN         156          CNY        NA
## 3 1392730000         CHN          CN         156          CNY        NA
## 4 1392730000         CHN          CN         156          CNY        NA
## 5 1392730000         CHN          CN         156          CNY        NA
## 6 1392730000         CHN          CN         156          CNY        NA
##   key_google_mobility key_apple_mobility key_jhu_csse key_nuts key_gadm
## 1                                                  CN       NA      CHN
## 2                                                  CN       NA      CHN
## 3                                                  CN       NA      CHN
## 4                                                  CN       NA      CHN
## 5                                                  CN       NA      CHN
## 6                                                  CN       NA      CHN
```

```
covid$date <- as.Date(covid$date)
```

```
covid <- covid %>%
  as_tsibble(key = "id", index = "date")
head(covid)
```

```
## # A tsibble: 6 x 47 [1D]
## # Key:       id [1]
##   id    date       confirmed deaths recove~1 tests vacci~2 peopl~3 peopl~4  hosp
##   <chr> <date>         <int>  <int>    <int> <dbl>   <dbl>   <int>   <int> <int>
## 1 China 2020-01-03         0      0       NA    NA      NA      NA      NA    NA
## 2 China 2020-01-04         1      0       NA    NA      NA      NA      NA    NA
## 3 China 2020-01-05         1      0       NA    NA      NA      NA      NA    NA
## 4 China 2020-01-06         4      0       NA    NA      NA      NA      NA    NA
## 5 China 2020-01-07         4      0       NA    NA      NA      NA      NA    NA
## 6 China 2020-01-08         4      0       NA    NA      NA      NA      NA    NA
## # ... with 37 more variables: icu <int>, vent <int>, school_closing <int>,
## #   workplace_closing <int>, cancel_events <int>,
## #   gatherings_restrictions <int>, transport_closing <int>,
## #   stay_home_restrictions <int>, internal_movement_restrictions <int>,
## #   international_movement_restrictions <int>, information_campaigns <int>,
## #   testing_policy <int>, contact_tracing <int>, facial_coverings <int>,
## #   vaccination_policy <int>, elderly_people_protection <int>, ...
```

```
class(covid)
```

```
## [1] "tbl_ts"     "tbl_df"     "tbl"        "data.frame"
```

```
covid <- covid %>%
  dplyr::select(id, date, confirmed)
```

```
covid <- covid %>%
  drop_na()
```

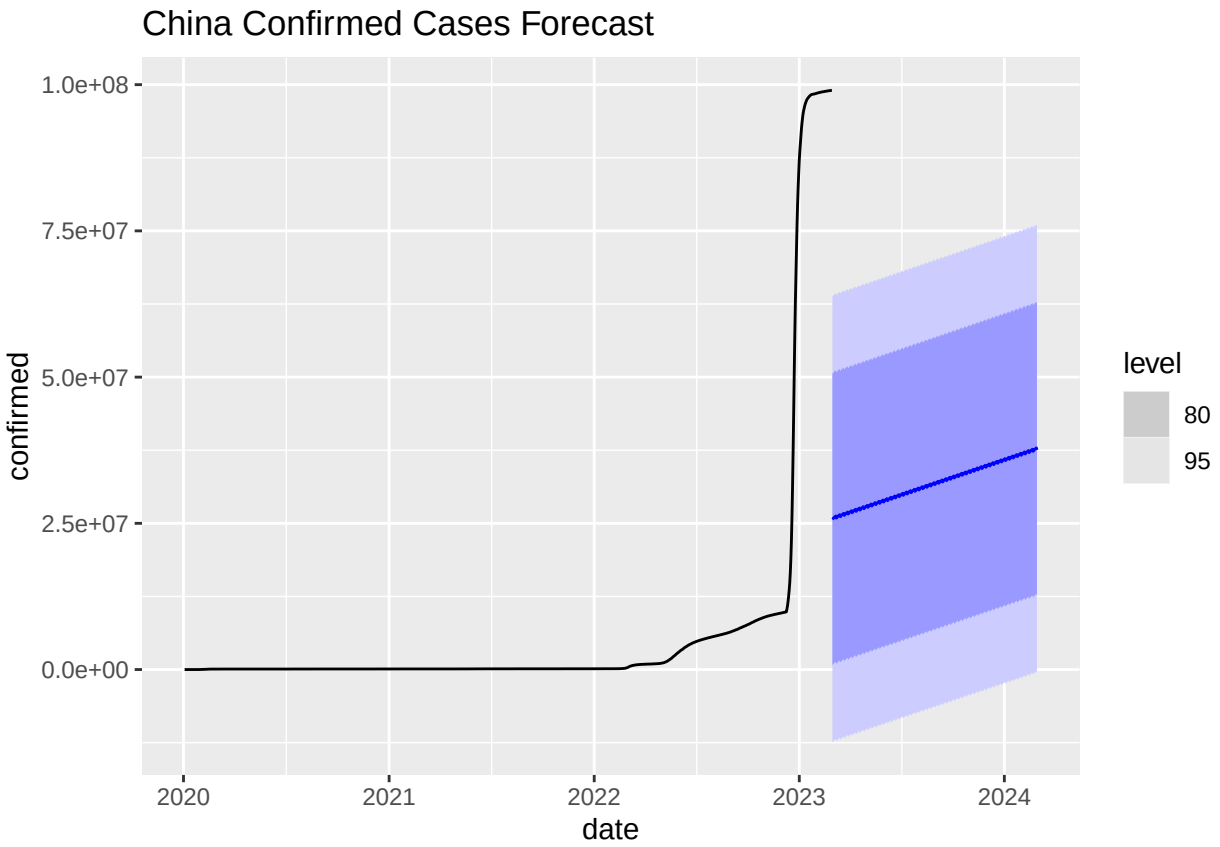
Time Series Linear Model

China

```
china <- covid %>%
  filter(id == "China") %>%
  select(date, confirmed)

head(china)
```

```
## # A tsibble: 6 x 2 [1D]
##   date       confirmed
##   <date>         <int>
## 1 2020-01-03         0
## 2 2020-01-04         1
## 3 2020-01-05         1
## 4 2020-01-06         4
## 5 2020-01-07         4
## 6 2020-01-08         4
```

```
china_fit <- china %>%
  model(TSLM(confirmed ~ trend() + season()))

china_fit %>%
  forecast(h = 365) %>%
  autoplot(china) +
  labs(title = "China Confirmed Cases Forecast")
```



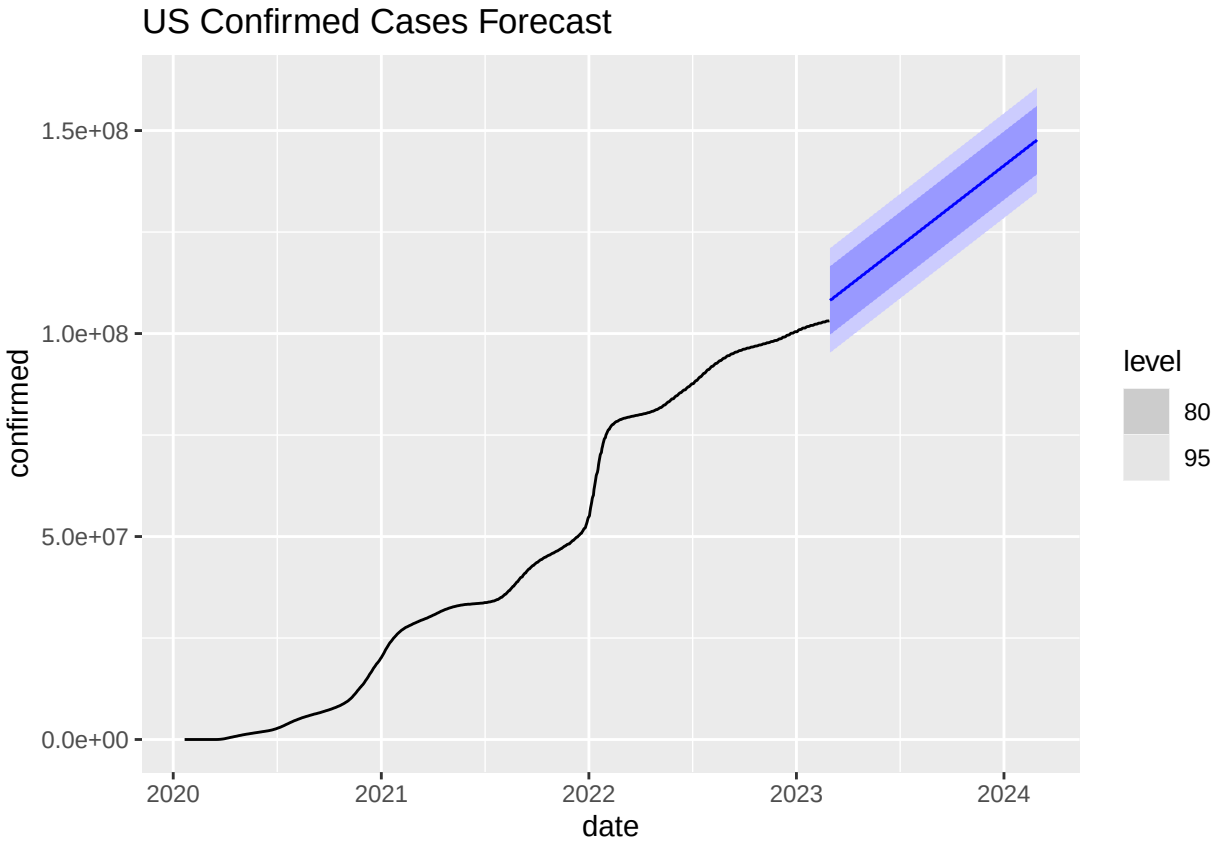
United States

```
us <- covid %>%
  filter(id == "United States") %>%
  select(date, confirmed)

head(us)
```

```
## # A tsibble: 6 x 2 [1D]
##   date       confirmed
##   <date>         <int>
## 1 2020-01-21         1
## 2 2020-01-22         1
## 3 2020-01-23         1
## 4 2020-01-24         2
## 5 2020-01-25         3
## 6 2020-01-26         5
```

```
us_fit <- us %>%
  model(TSLM(confirmed ~ trend() + season()))

us_fit %>%
  forecast(h = 365) %>%
```

```
  autoplot(us) +
  labs(title = "US Confirmed Cases Forecast")
```



Exponential Smoothing

China

Best Model

According to the RSME value, the appropriate exponential model is the holts method.

```
registerDoParallel(cores = 16)
china %>%
  stretch_tsibble() %>%
  model(dholts = ETS(confirmed ~ error("A") + trend("Ad") + season("N")),
        holts = ETS(confirmed ~ error("A") + trend("A") + season("N")),
        ses = ETS(confirmed ~ error("A") + trend("N") + season("N"))) %>%
  forecast(h = 365) %>%
  accuracy(china)
```

```
## Warning: 6 errors (2 unique) encountered for dholts
## [5] Not enough data to estimate this ETS model.
## [1] only 1 case, but 2 variables
```

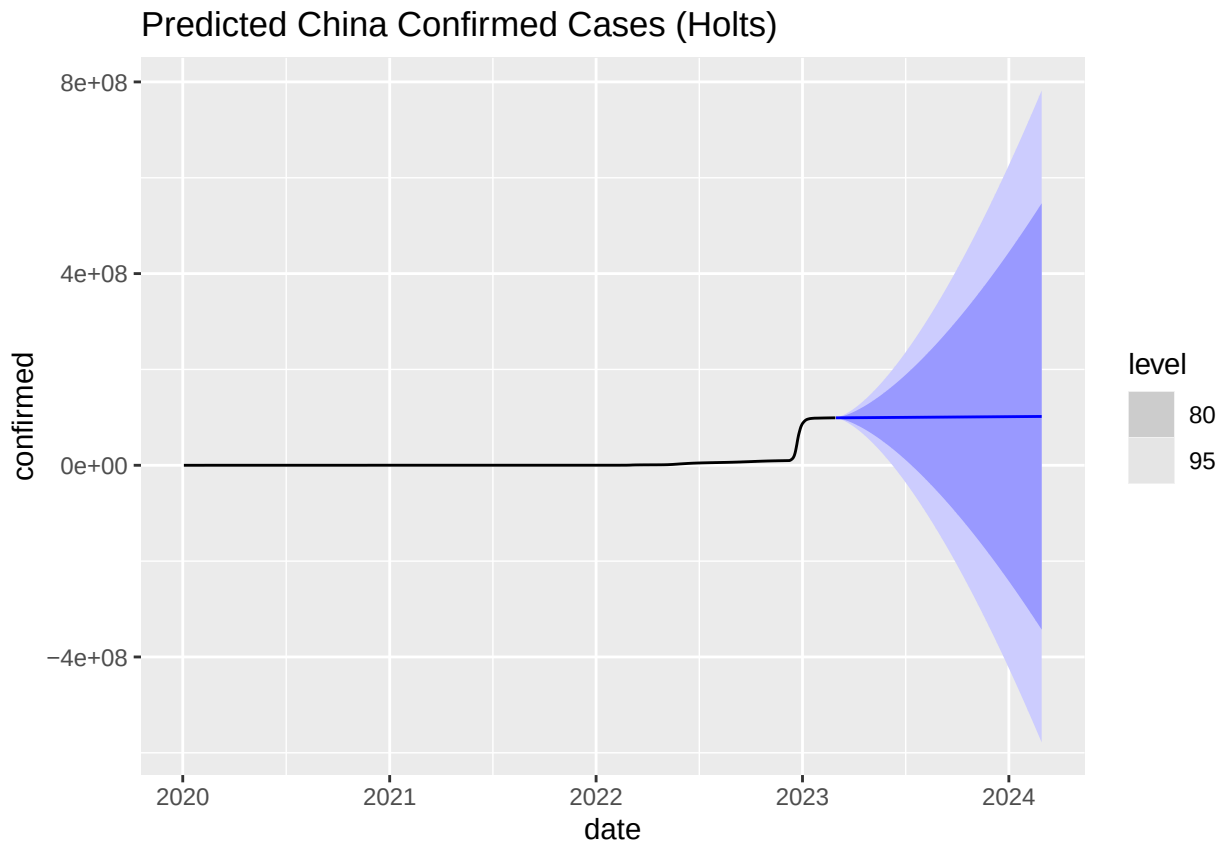
```
## Warning: 5 errors (2 unique) encountered for holts
## [4] Not enough data to estimate this ETS model.
## [1] only 1 case, but 2 variables
```

```
## Warning: 3 errors (1 unique) encountered for ses
## [3] Not enough data to estimate this ETS model.
```

```
## Warning: The future dataset is incomplete, incomplete out-of-sample data will be treated as missing.
## 365 observations are missing between 2023-03-01 and 2024-02-28
```

```
## # A tibble: 3 x 10
##   .model .type       ME      RMSE      MAE   MPE  MAPE  MASE RMSSE  ACF1
##   <chr>  <chr>    <dbl>     <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 dholts Test  6433955. 22671937. 6849061. 32.8   38.3  11.3  5.94 0.991
## 2 holts  Test  5690838. 22618077. 6637613.  9.96  49.3  11.0  5.92 0.990
## 3 ses    Test  7000446. 22793917. 7000446. 40.3   40.3  11.6  5.97 0.992
```

```
china %>%
  model(holts = ETS(confirmed ~ error("A") + trend("A") + season("N"))) %>%
  forecast(h = 365) %>%
  autoplot(china) +
  labs(title = "Predicted China Confirmed Cases (Holts)")
```



US

Best Model

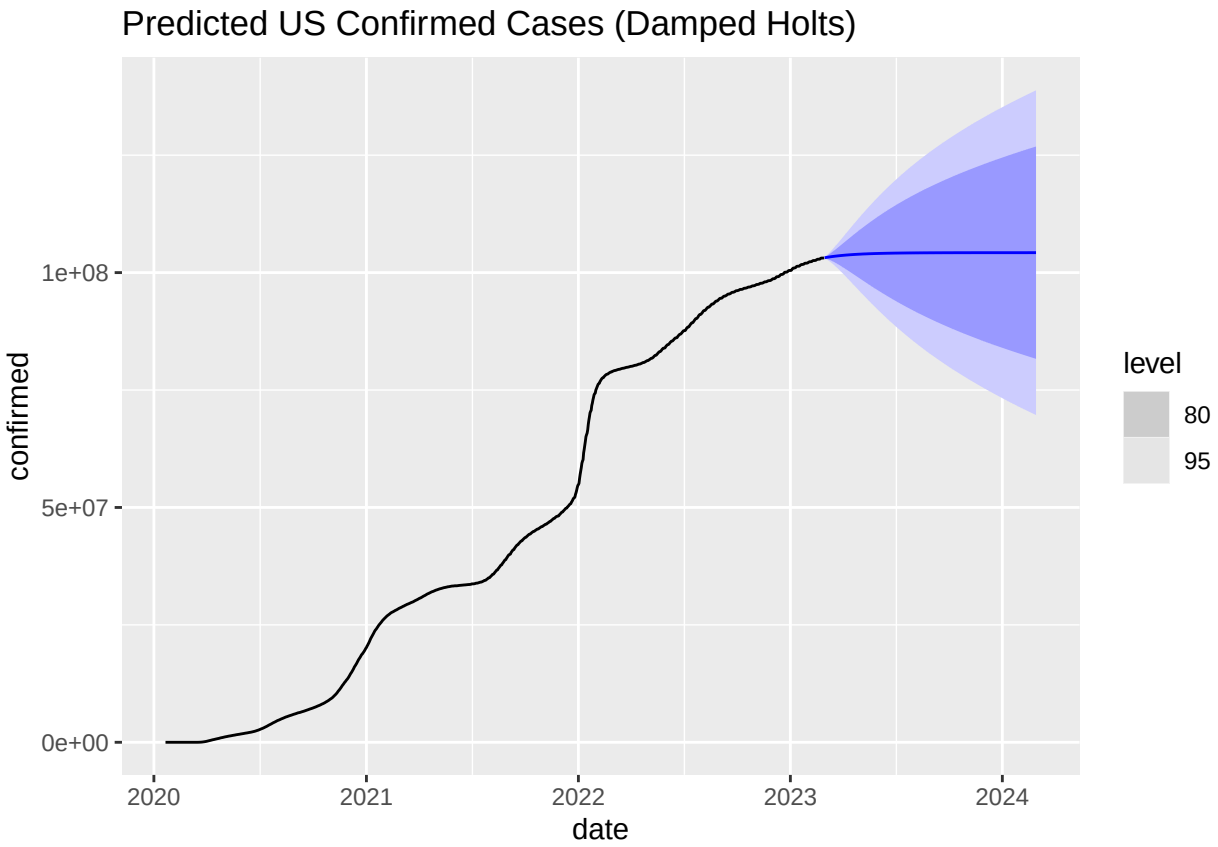
According to the RSME values, the appropriate exponential model is the damped holts method.

```
us %>%
  stretch_tsibble(.init = 10) %>%
  model(dholts = ETS(confirmed ~ error("A") + trend("Ad") + season("N")),
        holts = ETS(confirmed ~ error("A") + trend("A") + season("N")),
        ses = ETS(confirmed ~ error("A") + trend("N") + season("N"))) %>%
  forecast(h = 365) %>%
  accuracy(us)
```

```
## Warning: The future dataset is incomplete, incomplete out-of-sample data will be treated as missing.
## 365 observations are missing between 2023-03-01 and 2024-02-28
```

```
## # A tibble: 3 x 10
##   .model .type        ME      RMSE       MAE   MPE  MAPE  MASE RMSSE  ACF1
##   <chr>  <chr>     <dbl>     <dbl>     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 dholts Test  14214508. 20253535. 14686189.  34.4  35.0  23.0  19.8 0.993
## 2 holts  Test   1131641. 26694216. 13556262.  10.8  29.7  21.2  26.1 0.997
## 3 ses    Test  18160490. 23345197. 18160490.  43.2  43.2  28.4  22.9 0.991
```

```
us %>%
  model(dholts = ETS(confirmed ~ error("A") + trend("Ad") + season("N"))) %>%
  forecast(h = 365) %>%
  autoplot(us) +
  labs(title = "Predicted US Confirmed Cases (Damped Holts)")
```


Predicted US Confirmed Cases (Damped Holts)
confirmed
1e+08
5e+07
0e+00
2020
2021
2022
2023
2024
date
level
80
95