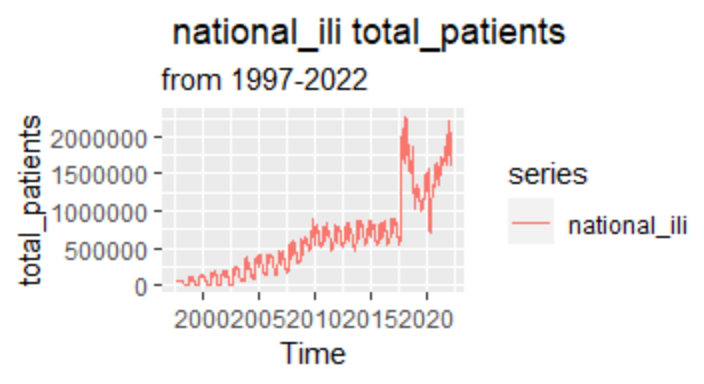
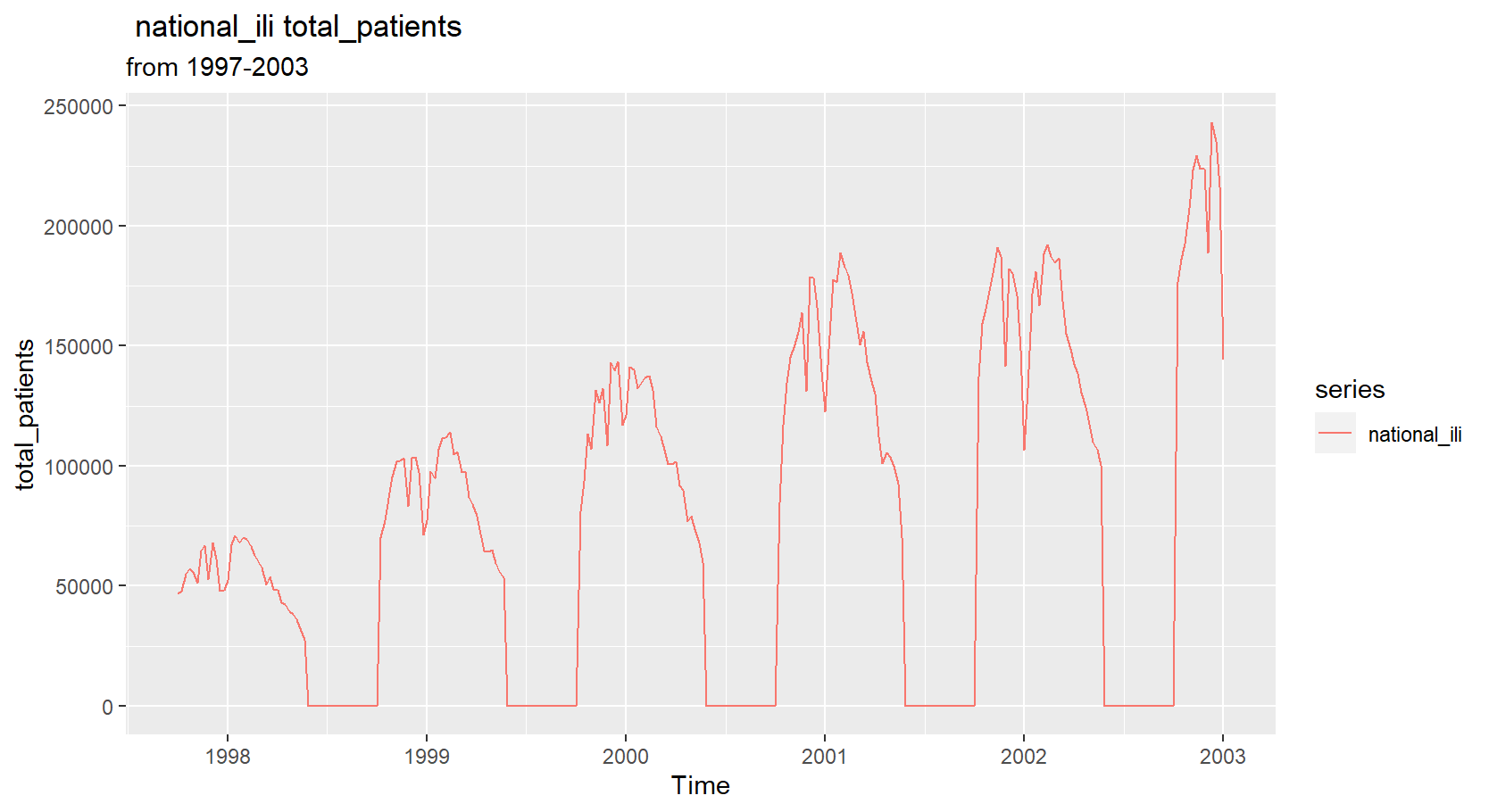
**מטלה בחיזוי באמצעות R**

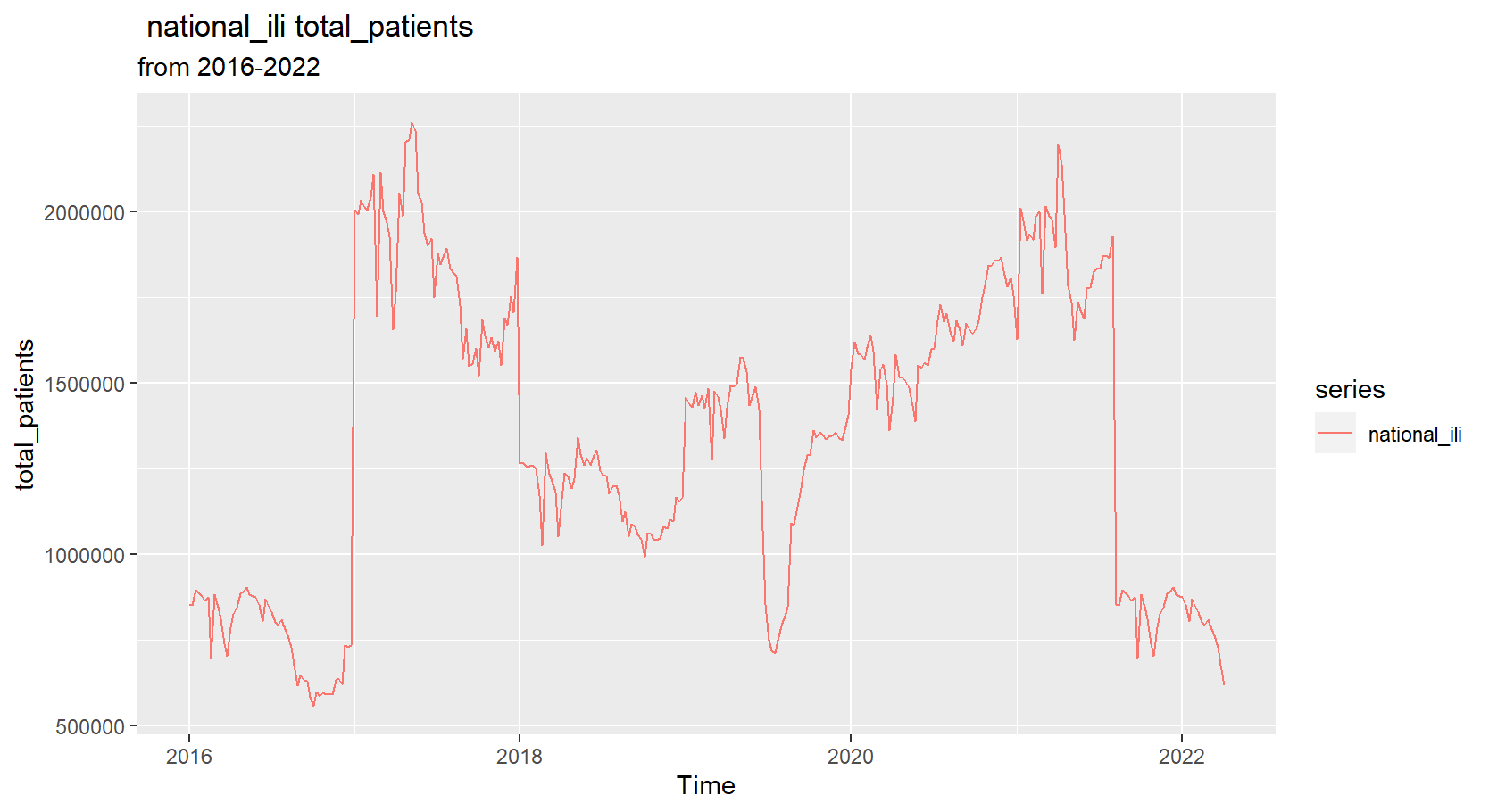
**סעיף ב'**

ראשית, על מנת להבין את הנתונים בצורה טובה יותר אנו ניצור גרפים שדרכם נוכל להבין איך הנתונים שלנו מתנהגים ואם יש נתונים חסרים או בעיות אחרות שאותן נצטרך לפתור. לצורך העיניין עשינו גרף כדי לראות כיצד מתנהגים הנתונים.

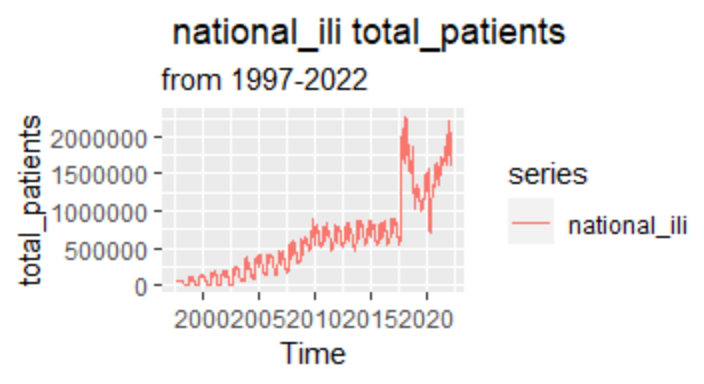


נשים לב לכמה בעיות במאגר הנתונים שלנו:

1. בין השנים 1997-2002 בשבועות 20-40 מצוין שאין חולי שפעת בכלל. הדבר נראה לנו לא הגיוני לכן אנו הסקנו שזה נובע ממחסור בנתונים. כשהסתכלנו על כל הנתונים לפי שנה ראינו שבאופן שיטתי בשבועות אלו יש ירידה משמעותית במספר חולי השפעת. אנו סבורים שזה קורה בגלל עונתיות, בחודשי הקיץ ישנה ירידה משמעותית במספר חולי השפעת. אפילו שנראה לנו תמוהה שיש 0 חולים בשנים אלה אנו נשאיר את הנתונים כמו שהם מכיוון שאנו נבצע חיזוי על פי התנהגות עונתית ובשנים אלו עדיין קיימת עונתיות ברורה שמתנהגת בצורה דומה גם לשנים שאחריה.
2. בחורף 2017-2018 נרשמה השפעת הקטלנית ביותר בעשור בארצות הברית, עם 61,000 מקרי מוות הקשורים בשפעת ו- 810,000 מקרי אשפוז הקשורים בשפעת. ניתן לראות זאת גם במאגר הנתונים שלנו בצורה מובהקת ששם היית עליה חדה בכמות החולים בשפעת.

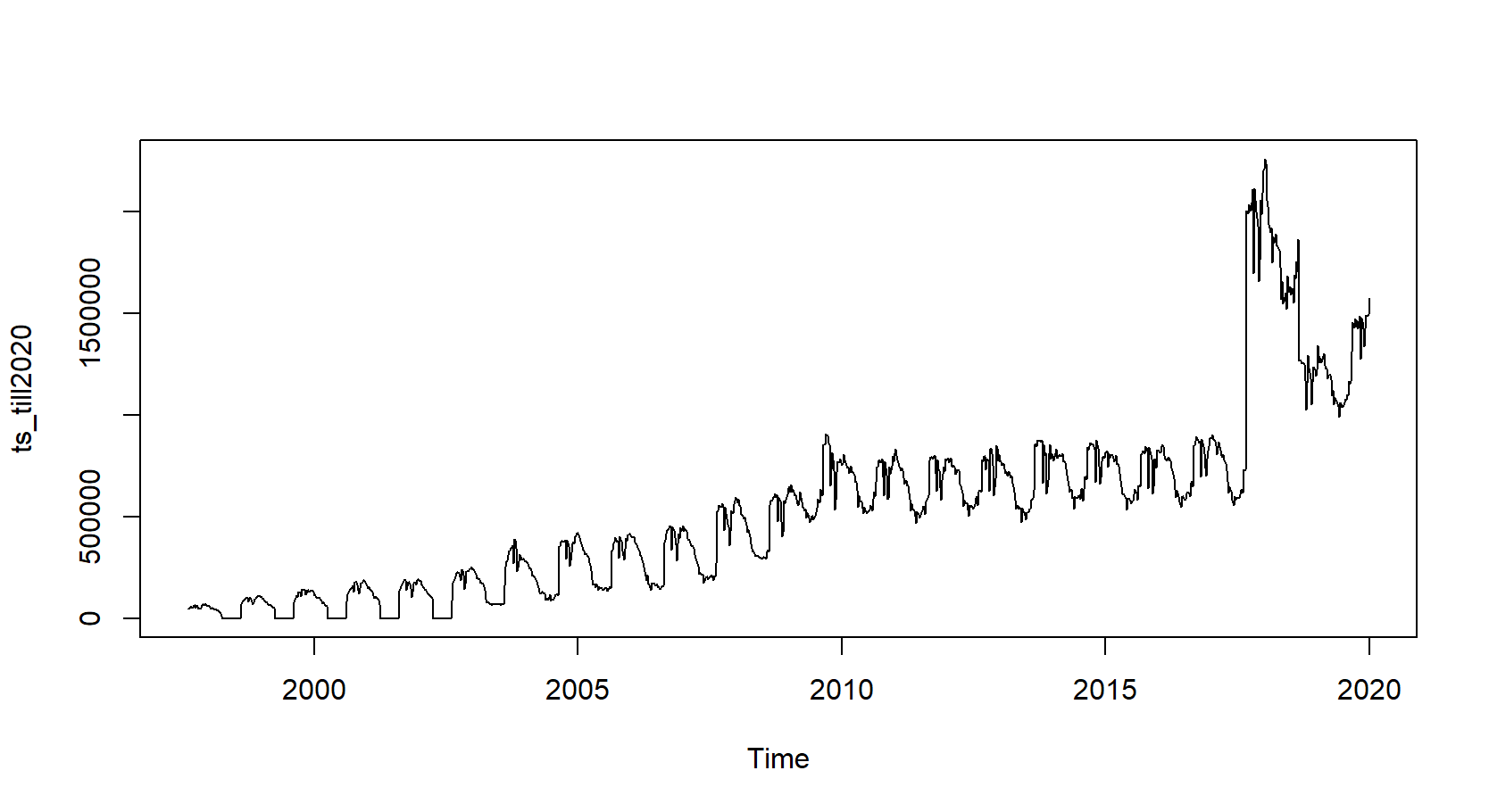


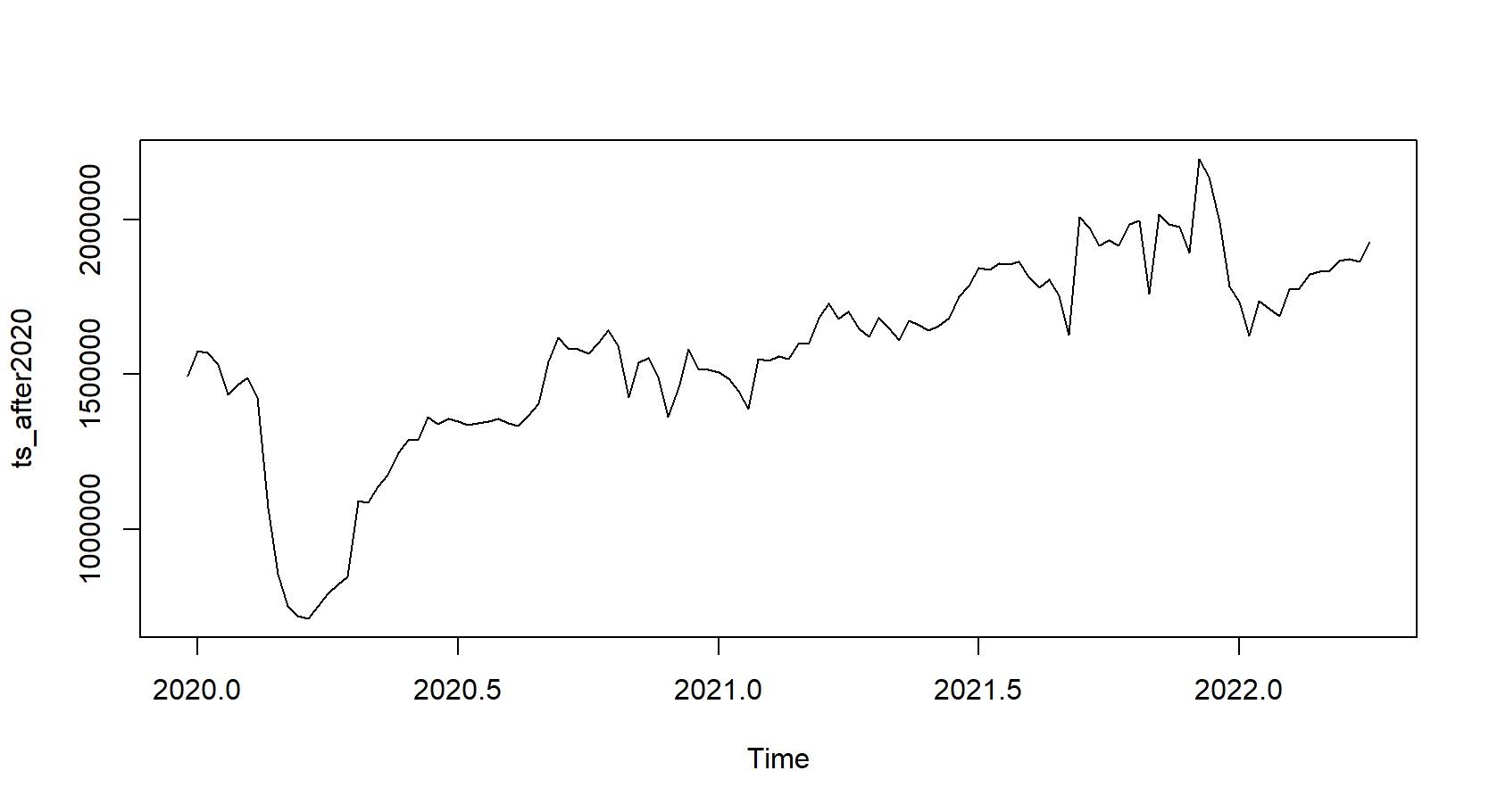
1. ניתן לראות עליה קבועה במספר חולי השפעת לאורך השנים, בכל שנה מספר החולים עלה פרט לשנת 2018 שבה היית התפרצות חריגה.



**סעיף ג'**

בשלב זה אנו מפצלים את מאגר הנתונים שלנו ל2 חלקים: train and test. נהוג להקצות 80 % ממאגר הנתונים לטובת אימון(train ) ו20 % לטובת מבחן(test). לכן שנינו במעט את טווח החלוקה של מאגר הנתונים. ללפני 2019 ואחרי 2019.

**מאגר הנתונים שישמש אותנו לאימון :**

**מאגר הנתונים שישמש אותנו למבחן :**

RMSE שקיבלנו מהשוואה לממוצע נע אל מול הדאטה ששייכת למבחן: 347633.7

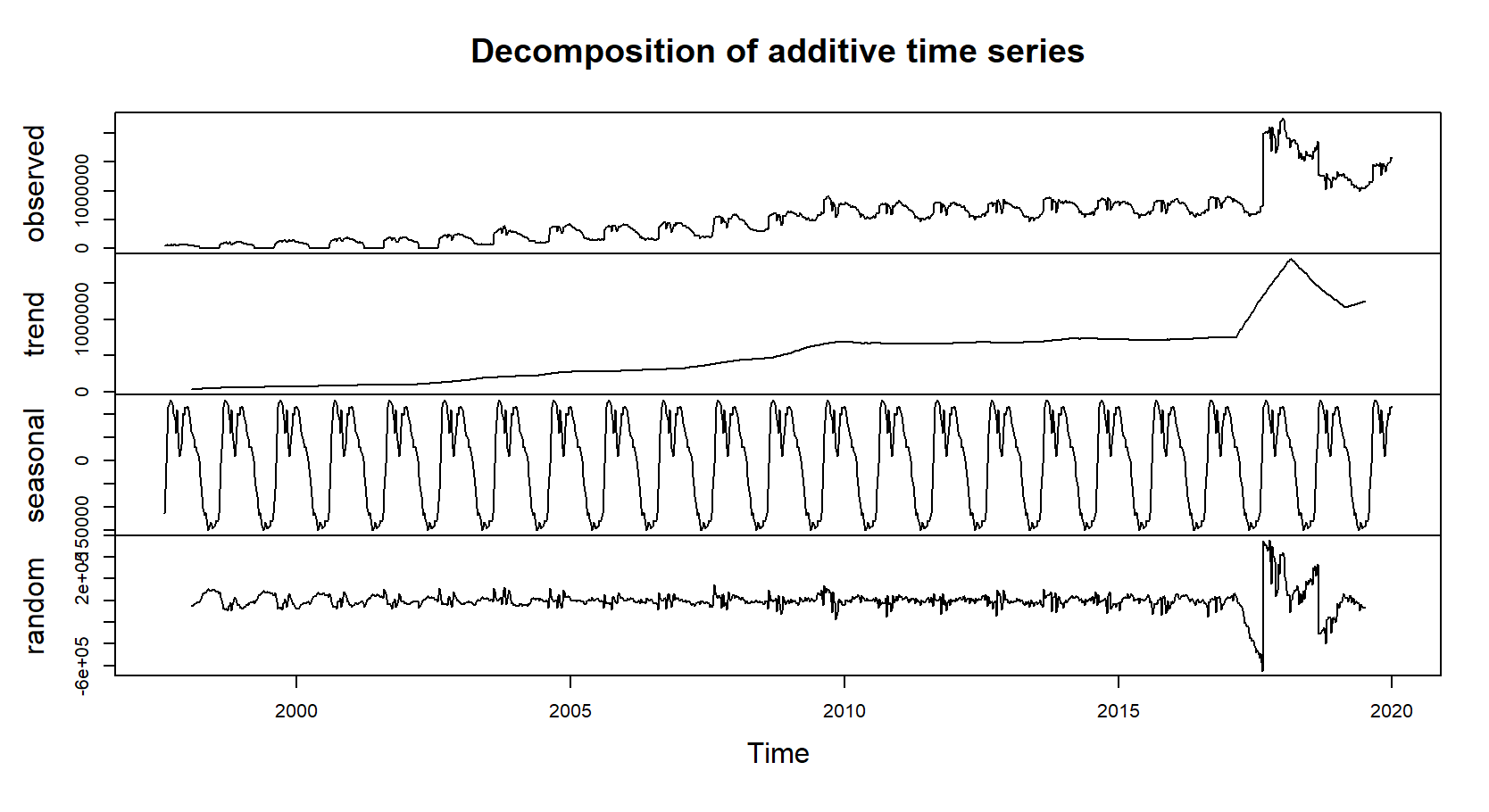
סעיף ד'

כפי שציינו מקודם ניתן לראות באופן ברור עונתיות ומגמת עליה בתחלואה. לכן, על מנת לבצע את החיזוי הטוב ביותר לנתונים הללו נשתמש במודל וינטרס לחיזוי שכולל מגמה + עונתיות. לטובת החיזוי נשתמש בכלים ממוחשבים(R).

בשלב הראשוני רצינו לבצע דה – קומפוזיציה על מנת לראות את מרכבי הדאטה שלנו.

components\_dfts <- decompose( ts\_till2020)

plot(components\_dfts)



גם פה ניתן להבחין בעונתיות ברורה, מגמה עולה וברעש גבוה בשנת 2018. נשים לב שהרעש יקשה עלינו ליצור מודל אמין ונצטרך למצוא דרך להתמודד עם זה בהמשך.

לאחר מכן הגדרנו מודל הולטוינטר על הדאטה שלנו ובדקנו את ערכי alpha ,beta , gamma . (בגלל שיש לנו 0 בנתונים השתמשנו ב "additive" ולא ב "multiplicative".)

HoltWinters1<- HoltWinters(ts.mydata, seasonal = "additive") # Creating a Holt Winters model

HoltWinters1$alpha # Displays the alpha value

## alpha

## 1

HoltWinters1$beta # Displays the beta value

## beta

## 0

HoltWinters1$gamma # Displays the gamma value

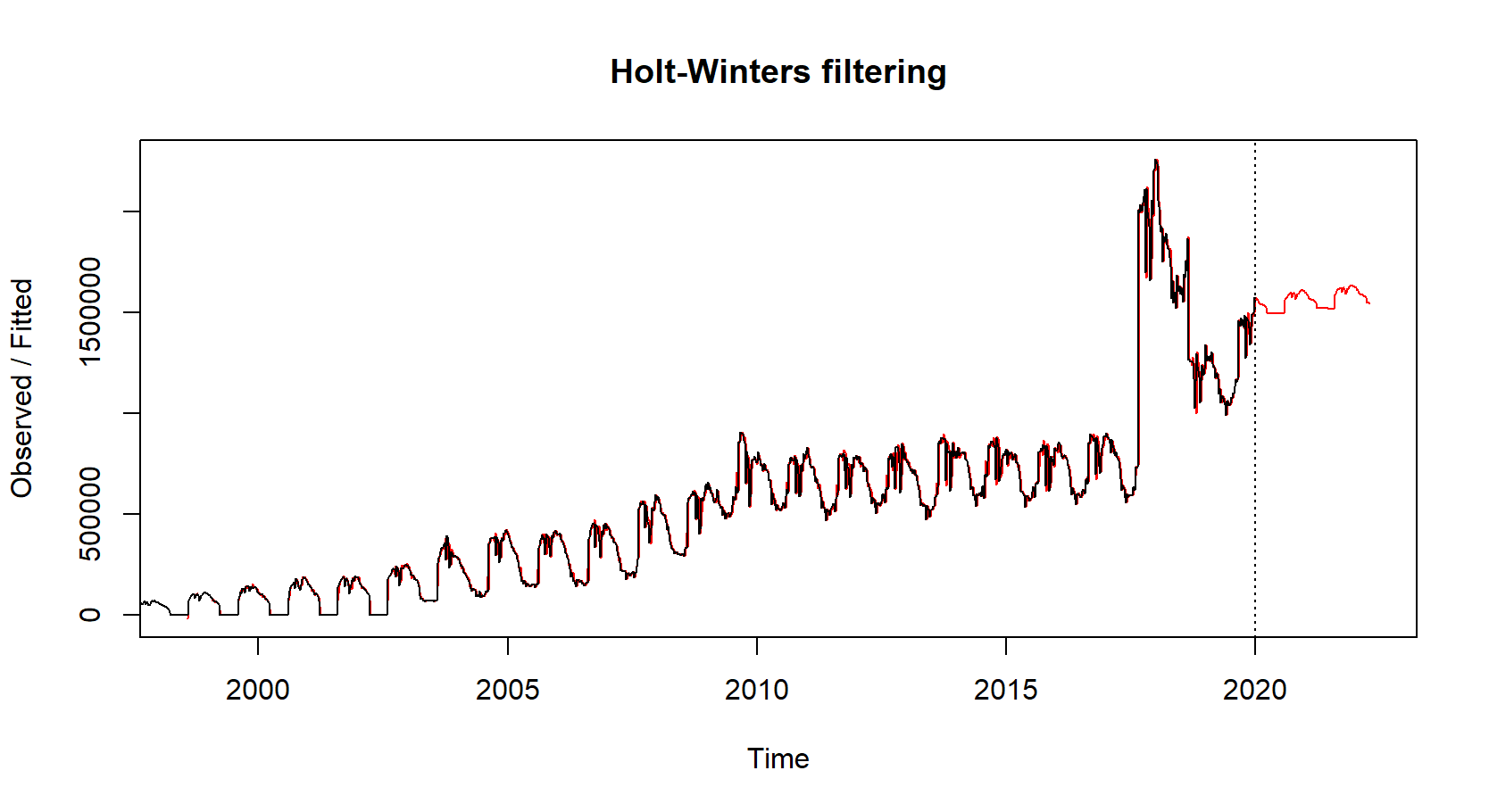
## gamma

## 1

HoltWinters1$SSE

accuracy( ts\_after2020 , Prediction )

כעת, רצינו לבדוק שערכי המקדמים שקיבלנו באמת מהווים אמצעי חיזוי טוב ואמין. לשם כך יצרנו גרף שיראה לנו את התחזית ל119 תקופות קדימה.(כדי שנוכל להשוות בין החיזוי של train לתוצאות האמתיות של test )



מכיוון שלא הכנסנו ערכים ל alpha ,beta , gamma והתוכנה בחרה אותם כערכים אופטימליים. ניתן לראות שהתחזית מאוד אמינה לנתונים המקוריים כשהם עוד קיימים אבל כשהיא צריכה לנבא נתונים שאינם קיימים הגרף תחזית לא מנבא בצורה טובה. בשלב זה בדקנו גם את ערך הSSE (סכום הטעויות בריבוע של המודל) וקיבלנו:

**RMSE= 292794.5 SSE =5.462368e+12**

כדי ליצור מודל טוב יותר נרצה לשנות את ערכי

alpha ,beta , gamma

**-Alpha**

מייחס חשיבות לערכים עצמם כשאר 1 מתייחס לתקופות האחרונות ו0 לתקופות הראשונות. הערך של אלפא כרגע הוא 1. ונרצה אומנם לתת חשיבות לתקופות האחרונות אבל פחות מהותית לכן בחרנו ב 0.8.

**Beta -**

מייחס חשיבות לשיפוע שלנו עצמם כשאר 1 מתייחס לתקופות האחרונות ו0 לתקופות הראשונות. ערך הבטה שלנו הוא כרגע 0. אנו נרצה לעלות ערך זה במעט כיוון שנרצה לתת משקל לשיפוע בתקופות האחרונות. אולם לא נרצה לתת חשיבות גדולה מידי לשנים האחרונות כי העלייה בתחלואה מתונה לאורך רוב השנים ורק מ2018 יש עלייה משמעותית בכמות החולים. לכן בחרנו ב 0.015

**Gamma-** מייחס חשיבות לעונתיות שלנו כשאר 1 מתייחס לתקופות האחרונות ו0 לתקופות הראשונות. ערך הגמה שלנו הוא כרגע 1, נשאיר אותו כפי שהוא מכיוון שבדה – קומפוזיציה ניתן לראות שהעונתיות קבועה לאורך כל השנים.

HoltWinters1<- HoltWinters(ts\_till2020, seasonal = "additive",alpha = 0.8 ,beta = 0.015 , gamma = 1) # Creating a Holt Winters model

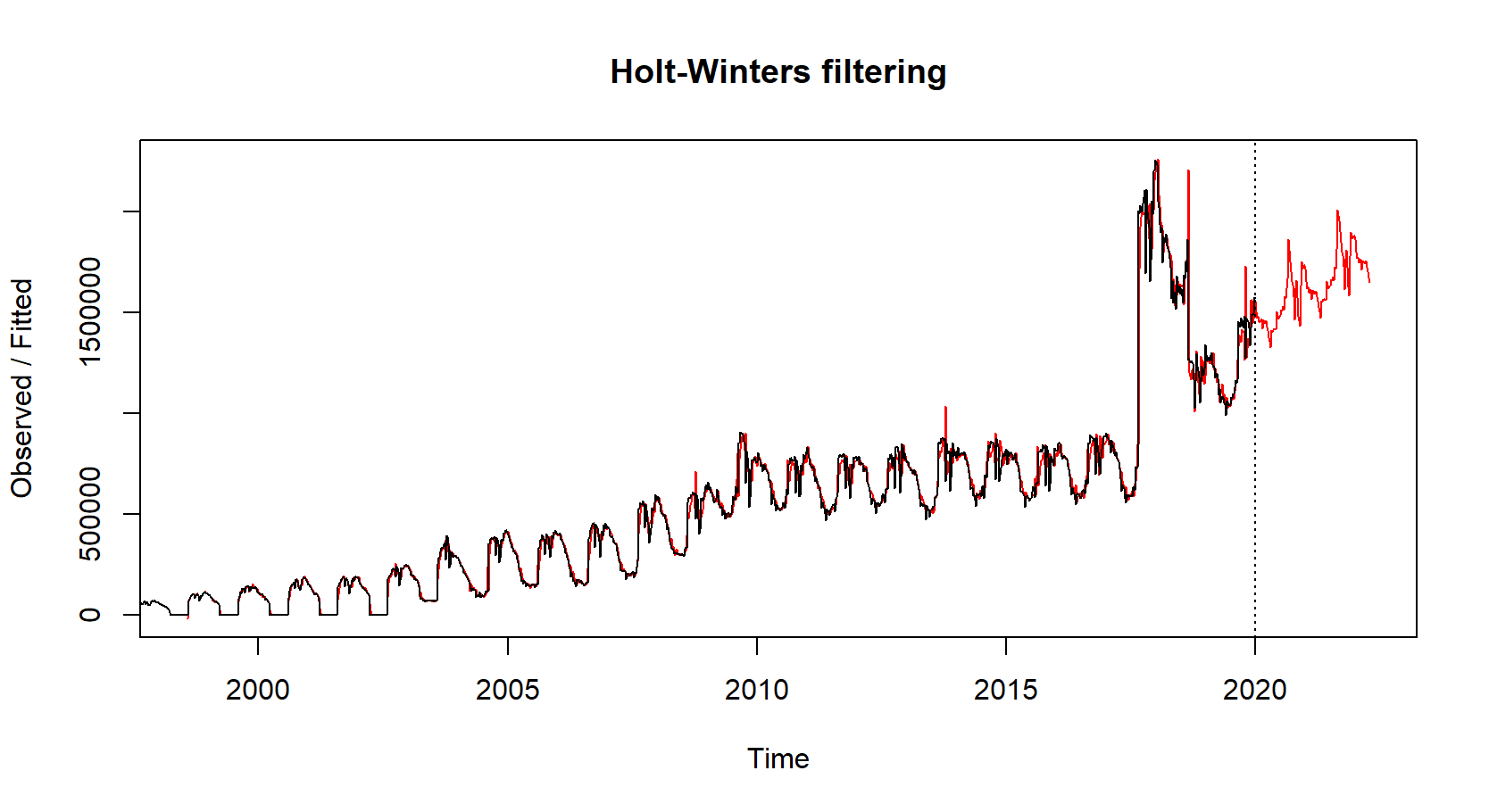
Prediction <- predict(HoltWinters1, 119, prediction.interval = FALSE) #Making predict for 300 periods ahead

plot(HoltWinters1 ,Prediction) # Print the forecast

HoltWinters1$SSE

accuracy( ts\_after2020 , Prediction )accuracy( ts\_after2020 , Prediction)

**RMSE= 238883.7 , SSE= 5.59263e+12**



**ניסינו לחזות גם על פי מודל הולטר וינטר עם עונתיות multiplicative (מכפלה בגורם העונתיות) לשם כך השתמשנו בנתונים רק אחרי 2003 היכן שכבר לא היו נתונים חסרים.**

ts\_after2003 <- window(ts.mydata , start=c( 2003,1 ) , end =c(2020,1))

HoltWinters2<- HoltWinters(ts\_after2003, seasonal = "multiplicative") # Creating a Holt Winters model

Prediction <- predict(HoltWinters2, 119, prediction.interval = FALSE) #Making predict for 119 periods ahead

plot(HoltWinters2, Prediction) # Print the forecast

HoltWinters2$SSE

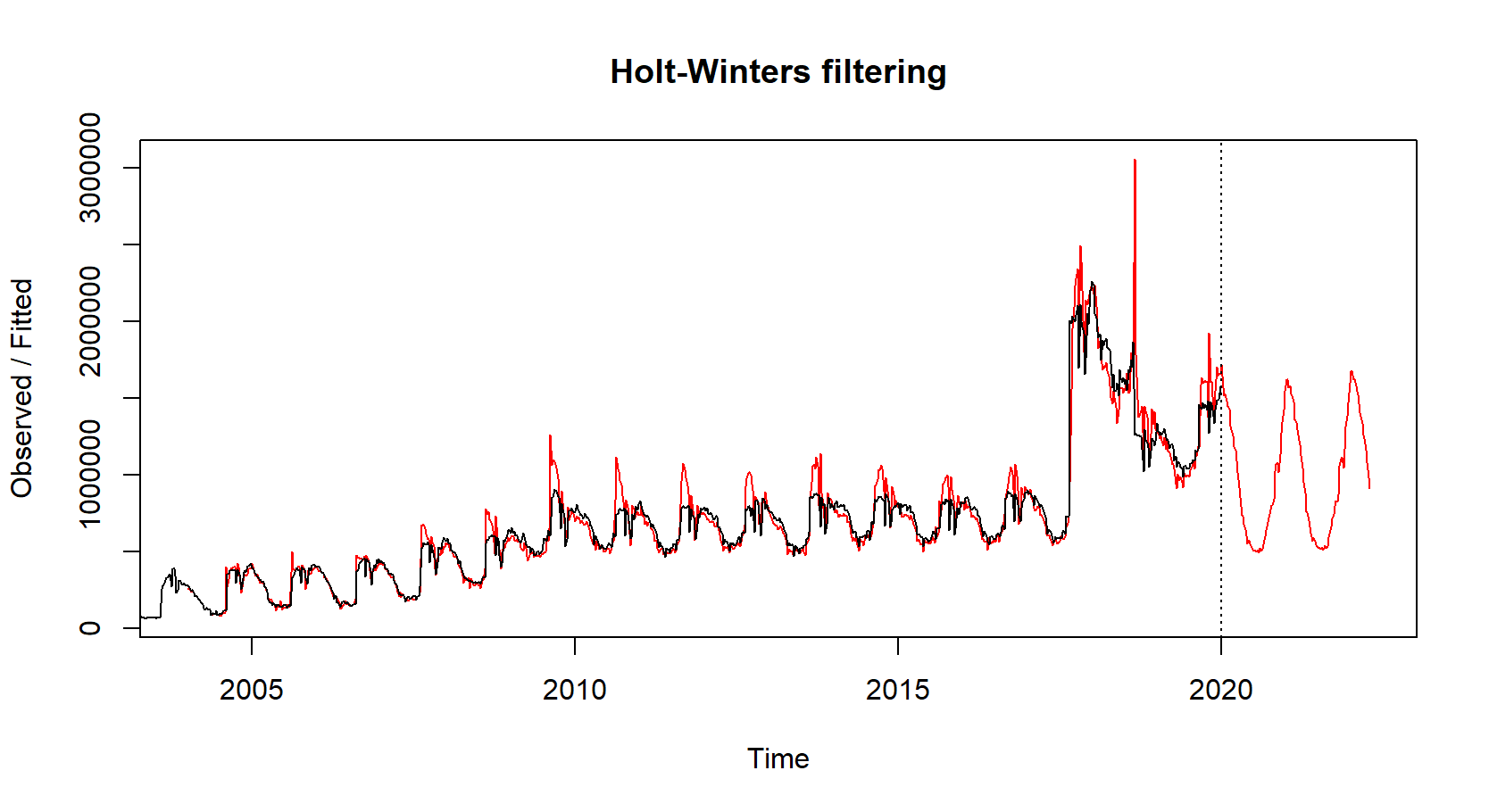
HoltWinters2$alpha#0.5528214

HoltWinters2$beta#0

HoltWinters2$gamma#1

accuracy( ts\_after2020 , Prediction )

**RMSE = 760025.8**



**לאחר ניסיון נוסף בוא שינינו את ערכי alpha beta gamma על פי השיקולים שרשמנו מקודם קיבלנו:**

#second prediction with multi

HoltWinters\_multi\_2<- HoltWinters(ts\_after2003, seasonal = "multiplicative",alpha = 0.7,beta = 0 ,gamma = 0.7) # Creating a Holt Winters model

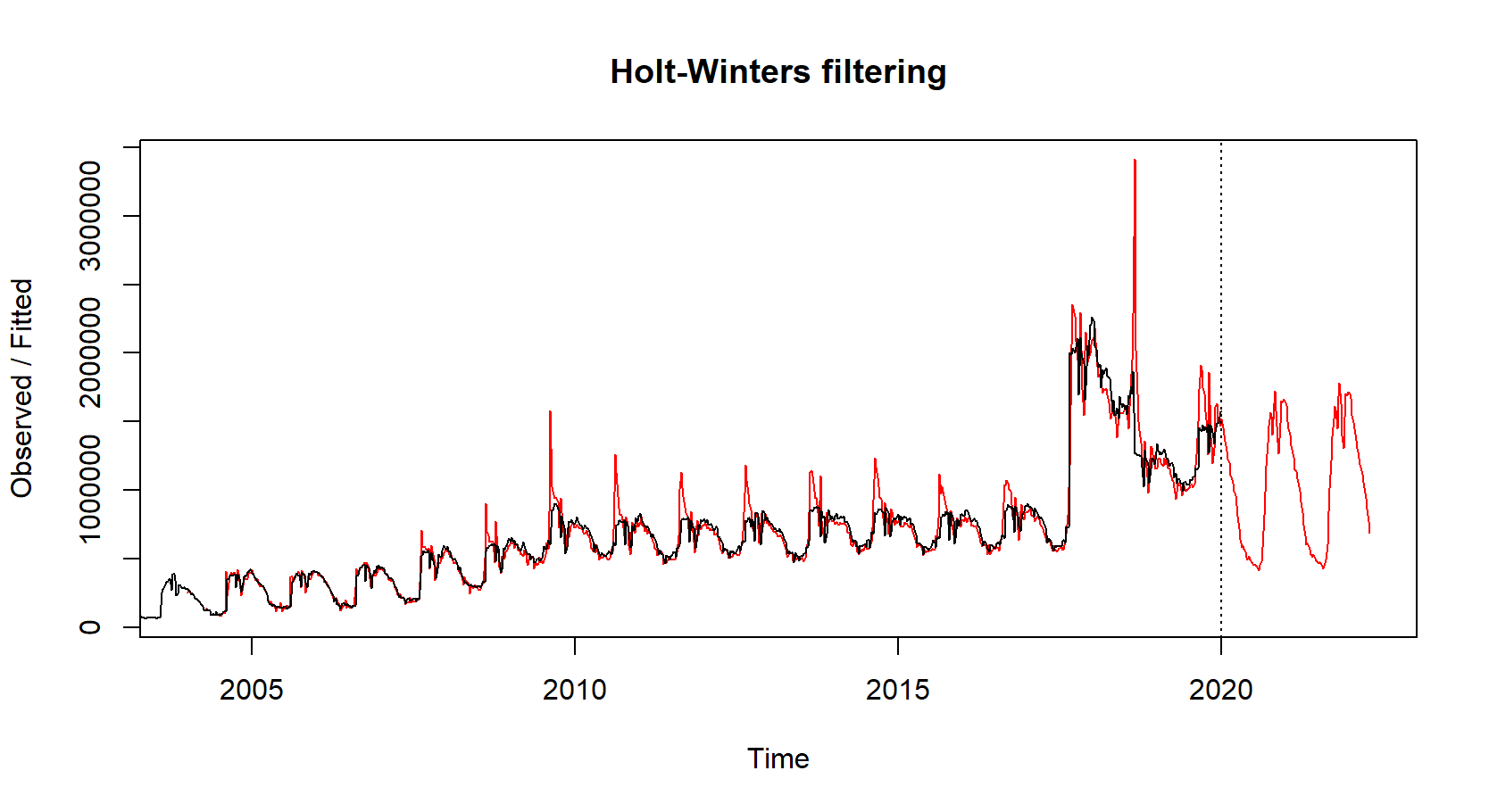
Prediction <- predict(HoltWinters\_multi\_2, 119, prediction.interval = FALSE) #Making predict for 500 periods ahead

plot(HoltWinters\_multi\_2, Prediction) # Print the forecast

HoltWinters\_multi\_2$SSE

accuracy( ts\_after2020 , Prediction )

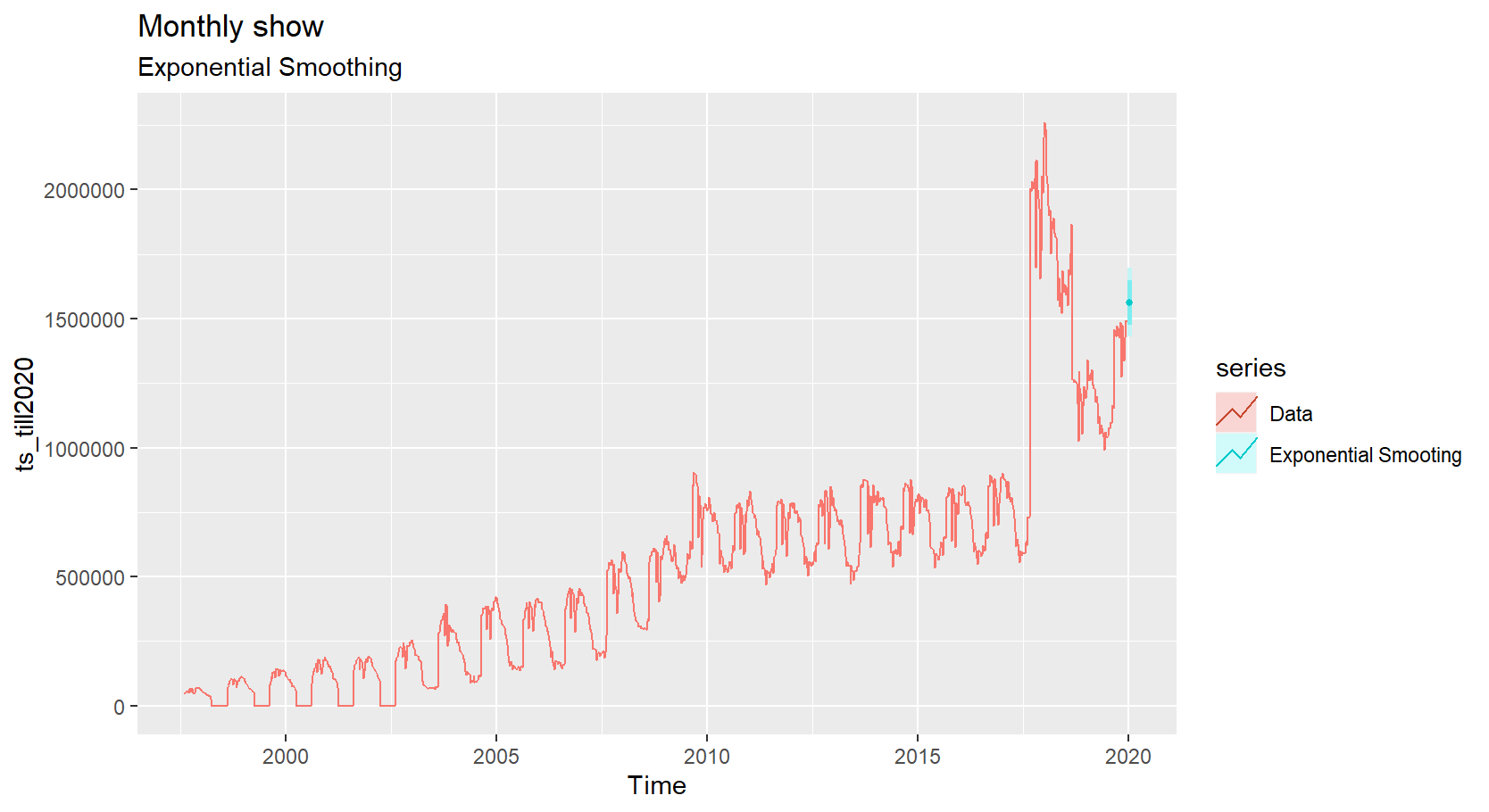
**RMSE = 700827.8**



בגרף זה ניתן לראות שהרעש גבוה והחיזוי נמוך מהצפייה שלנו. מבין 2 האפשרויות שבדקנו בהולט וינטרס נעדיף לשנות הערכים בעונתיות של additive מכיוון שהם אמינים יותר.(בעליRMSE קטן יותר)

**נבדוק אפשרת לחיזוי גם באמצעות החלקה מעריכית:**

עבור תקופה 1 נקבלת תוצאה אמינה :



אולם עבור מספר גבוה של תקופות נקבל חיזוי לא טוב:



**#predict with Exponential Smoothing**

fc <- ses(ts\_till2020, h=119)

p1<-autoplot(ts\_till2020, series="Data") +

autolayer(fc, series="Exponential Smooting") + labs(title = "Monthly show",subtitle ="Exponential Smoothing" )

p1

accuracy( ts\_after2020 , fc$mean )

**RMSE = 314465.2**

**אפשרת לחיזוי גם באמצעות ממוצע נע:**

בחרנו לא לנסות לחזות על פי ממוצע נע מכיוון שאנחנו מנסים ליצור חיזוי לטווח גדול מאוד של זמן.על פי התנהגות הנתונים שלנו אנחנו מבינים שמדובר במודל עונתי עם חריגות מסויימות בתקופות שונות כך שלדעתנו ליצור ממוצע נע לא יענה לנו על החיזוי המתאים לסוג מודל זה. בנוסף לכך אם נרצה לעשות חיזוי דרך ממוצע נע סביר להניח שנצליח לחזור בברור רק טווח קרוב מאוד אך לא נצליח לחזות בצורה טובה חודשים קדימה.

**אפשרות לחיזוי בעזרת ריגרסיה:**

#predict with regression

plot(ts\_till2020)

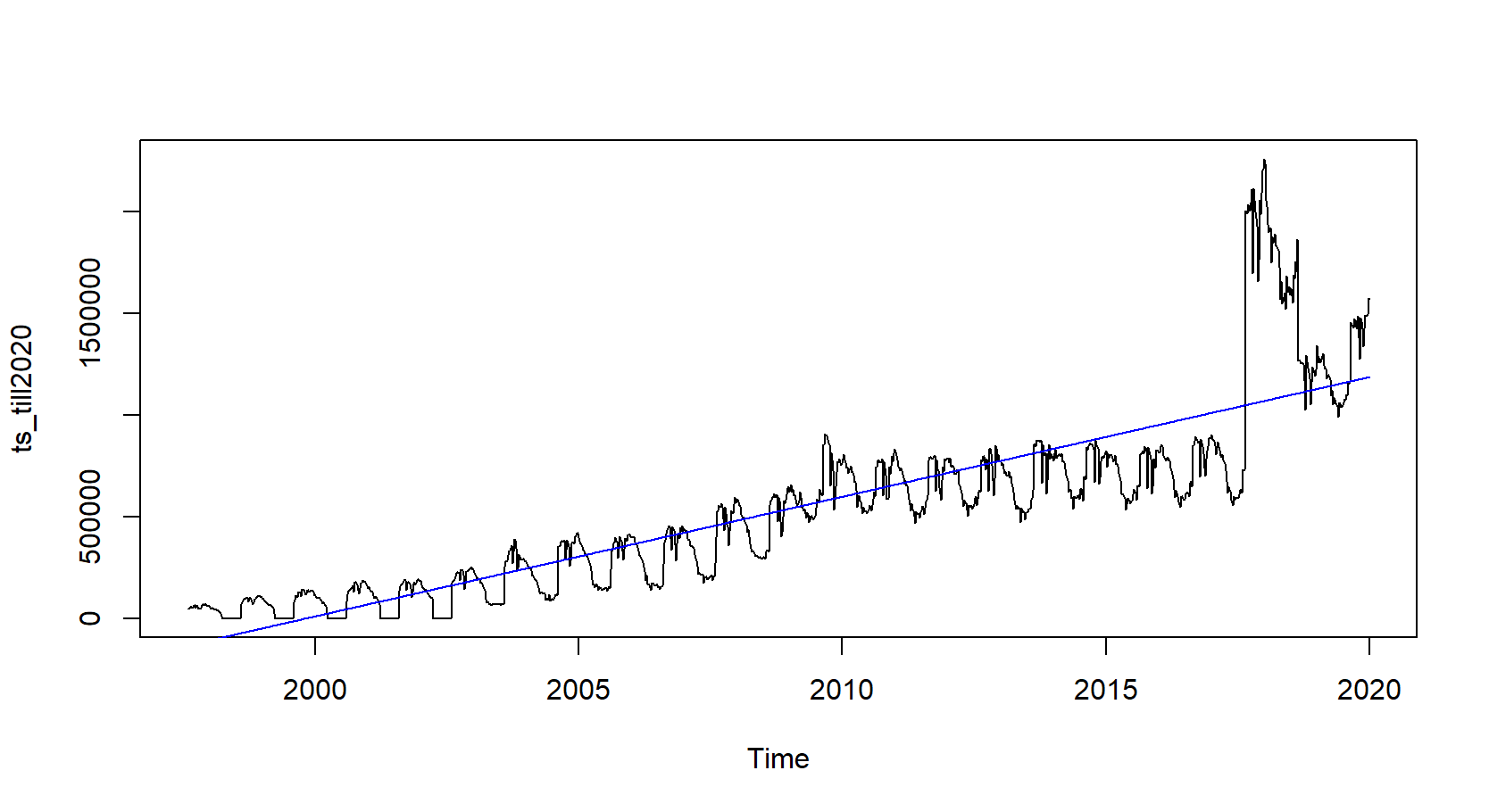
dat.lm1 <- tslm(ts\_till2020 ~ trend)

summary(dat.lm1)

lines(dat.lm1$fitted,col = "blue", lwd = 1)

accuracy(dat.lm1)

**rmse-** **225562.6**



**כעת, נבצע את מודל הולט וינטרס עם המקדמים שמצאנו מקודם על חודשים 5-10 בשנת 2022.**

HoltWinters\_flu<- HoltWinters(ts\_after2020, seasonal = "additive",alpha = 0.8 ,beta = 0.015 , gamma = 1) # Creating a Holt Winters model

HWflu.pred <- predict(HoltWinters\_flu, 29, prediction.interval = TRUE, level=0.95)#

