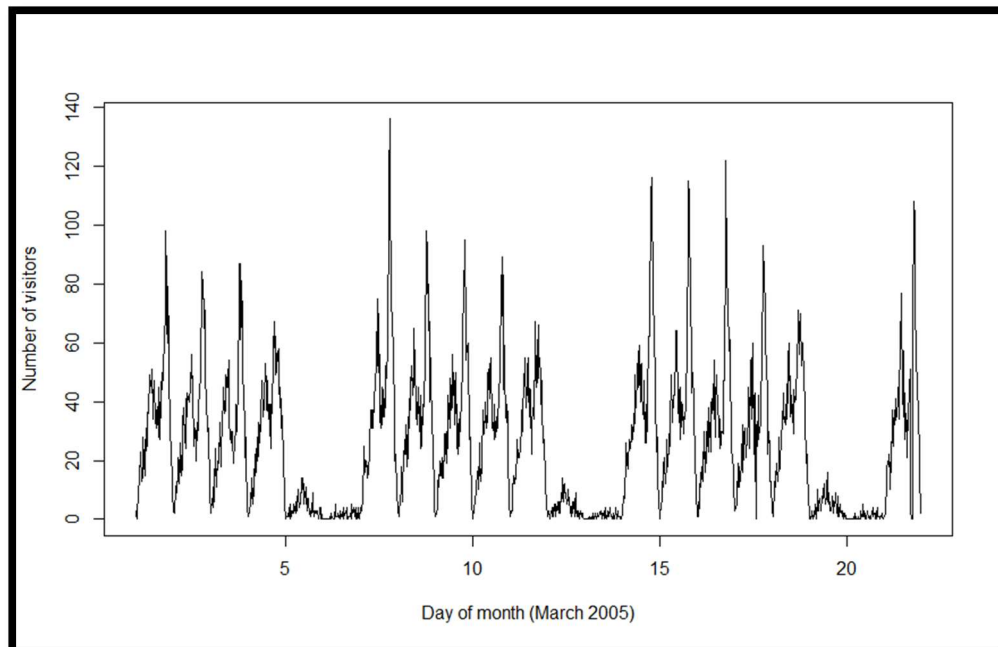


יישומי R בכריית נתונים, תרגיל 4 – קובץ פלט

2. פלט הגרף המתקבל מהרצת נתוני המידע ההיסטורי בלבד. לתאריכים 1.3.05-21.05.2005. כולל.



3. פירוט רכיבי הסדרה העתית, לפי גרף הפלט בסעיף 2.

א. **Level** - הסדרה מתחילה מערך 1 שהוא מאוד קרוב ל-0, יחד עם זאת לא מתחילה מ-0 מוחלט.

ב. **Trend** - לינארי או אקספוננציאלי? בחינה ויזואלית לא מעידה על מגמה אקספוננציאלית ו/או

לינארית. נבחן את המגמה על ידי התאמת ה-trend והשוואת ערכי ה-RMSE. פירוט בסעיף 4.

ג. **Seasonality** - לסדרה עונתיות מורכבת, אנחנו רואים מחזוריות פחות או יותר דומה בכל אחד מימות

השבוע, ואילו מחזוריות נפרדת בימים 6-7 של כל שבוע (סופי שבוע?). כלומר - פעם אחת יש

מחזוריות בכל יום ופעם שניה יש מחזוריות בכל שבוע.

ד. **Noise** - תצפית 428, מספר חריג של מבקרים בתאריך 7.3.05 בשעה 18:45 - 136 מבקרים. נוריד

תצפית זו בעזרת יצירת סדרה עתית אשר אינה מכילה אותה.

4. מהתאמת מודל רגרסיה ל-trend לינארי ואקספוננציאלי התקבלו אותם ערכי השגיאות בשני המודלים.
הרצת summary נתנה פלט שגיאה, קרי לא ניתן להצביע על trend מובהק, בשונה ממסד ה-cocacola.

```
> accuracy(bicup2.lm)
```

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	4.242199e-16	23.26915	18.96233	-Inf	Inf	0.9945283

```
> summary(bicup2.lm)
```

Call:
tslm(formula = bicup2.ts ~ trend)

Residuals:
Error in quantile.default(resid) :
missing values and NaN's not allowed if 'na.rm' is FALSE

תחזית ביקוש בשלושת הימים החסרים, לפי טבלת ה-summary, רואים ערכי trend ו-intercept.
ביצועי המודל (לינארי), לפי RMSE.

```
> summary(train.lm)
```

Call:
tslm(formula = train.ts ~ trend)

Residuals:
Min 1Q Median 3Q Max
-30.452 -21.417 -1.638 13.710 92.203

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 25.183317 1.368671 18.400 <2e-16 ***
trend 0.004646 0.002089 2.224 0.0263 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

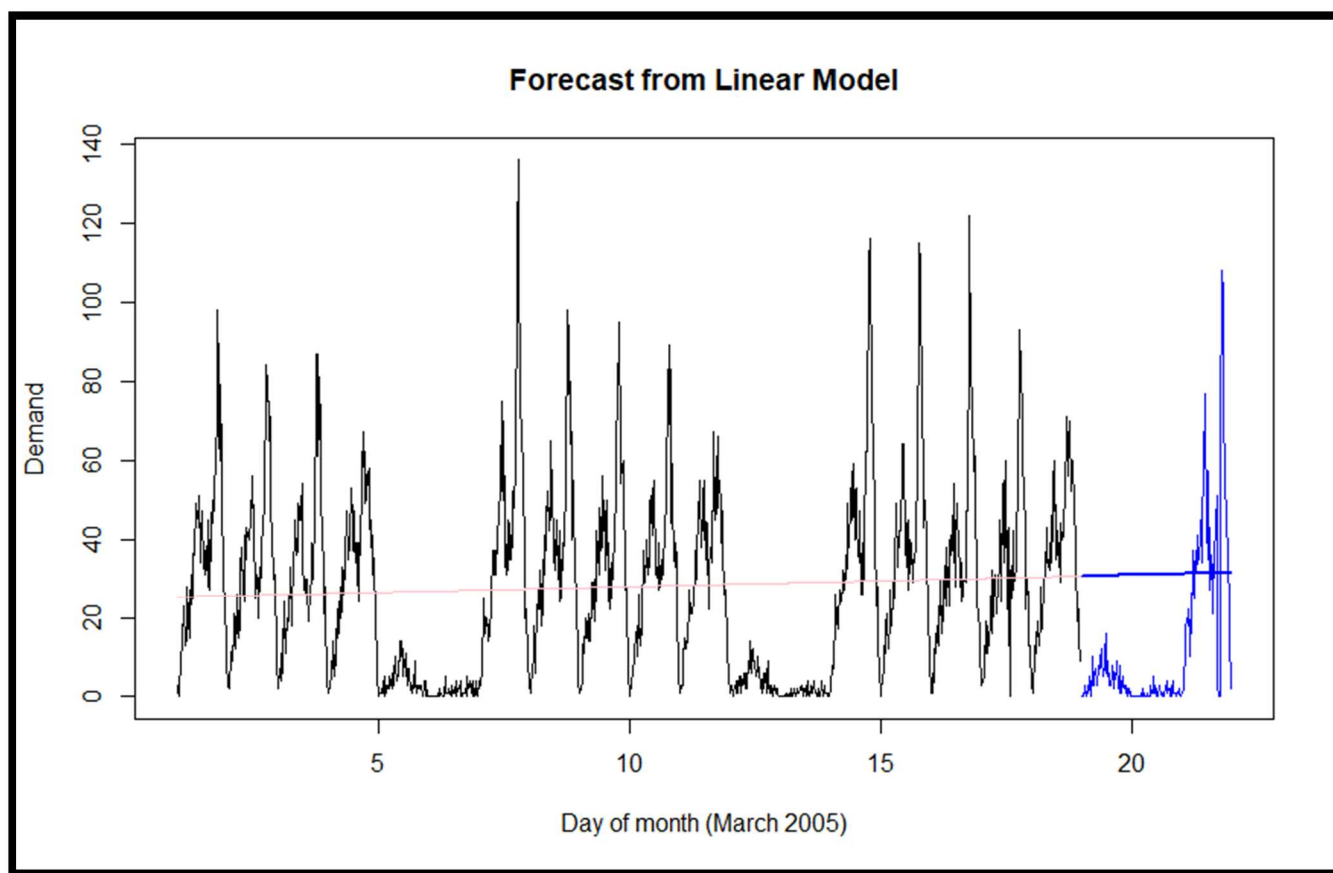
Residual standard error: 23.03 on 1132 degrees of freedom
Multiple R-squared: 0.004351, Adjusted R-squared: 0.003471
F-statistic: 4.947 on 1 and 1132 DF, p-value: 0.02634

```
> train.lm.pred <- forecast(train.lm, h = nvalid, level = 0)
> plot(train.lm.pred, ylab = "Demand", xlab = "Day of month (March 2005)", main = "Forecast from Linear Model")
> lines(train.lm$fitted, lwd = 0.5, col="pink")
> lines(valid.ts, col = "blue")
> accuracy(train.ts, train.lm$fitted) #training
```

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	1.98508e-16	23.00939	18.59133	-0.03722032	66.96824	0.9246427	4933.527

```
> accuracy(valid.ts, train.lm.pred$mean) #validation
```

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	17.35409	27.17868	24.91266	56.51225	80.71682	0.8999126	5830.135



5. התאמת עונתיות פשוטה למודל, נותנת את ערכי הפלט. התאריכים אותם אנחנו חוזים 22-24.5.05 לפי התחזית, הם ימי חול (בהם ההתנהגות "רגילה") ולכן עונתיות פשוטה יכולה לתת מענה לצורך המבוקש.

```

      ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set -6.038107e-17 17.26323 12.17225 -3.139791 48.62309 0.8919588 2.716607
> accuracy(valid.ts, train.lm.pred$mean) #validation
      ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 16.6527 27.53602 22.52339 55.53852 74.59477 0.9121773 5.336762
>

```

