

ΔΙΑΧΕΙΡΙΣΗ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ - ΕΡΓΑΣΤΗΡΙΟ

Χρονοσειρές (Time-Series) στην R

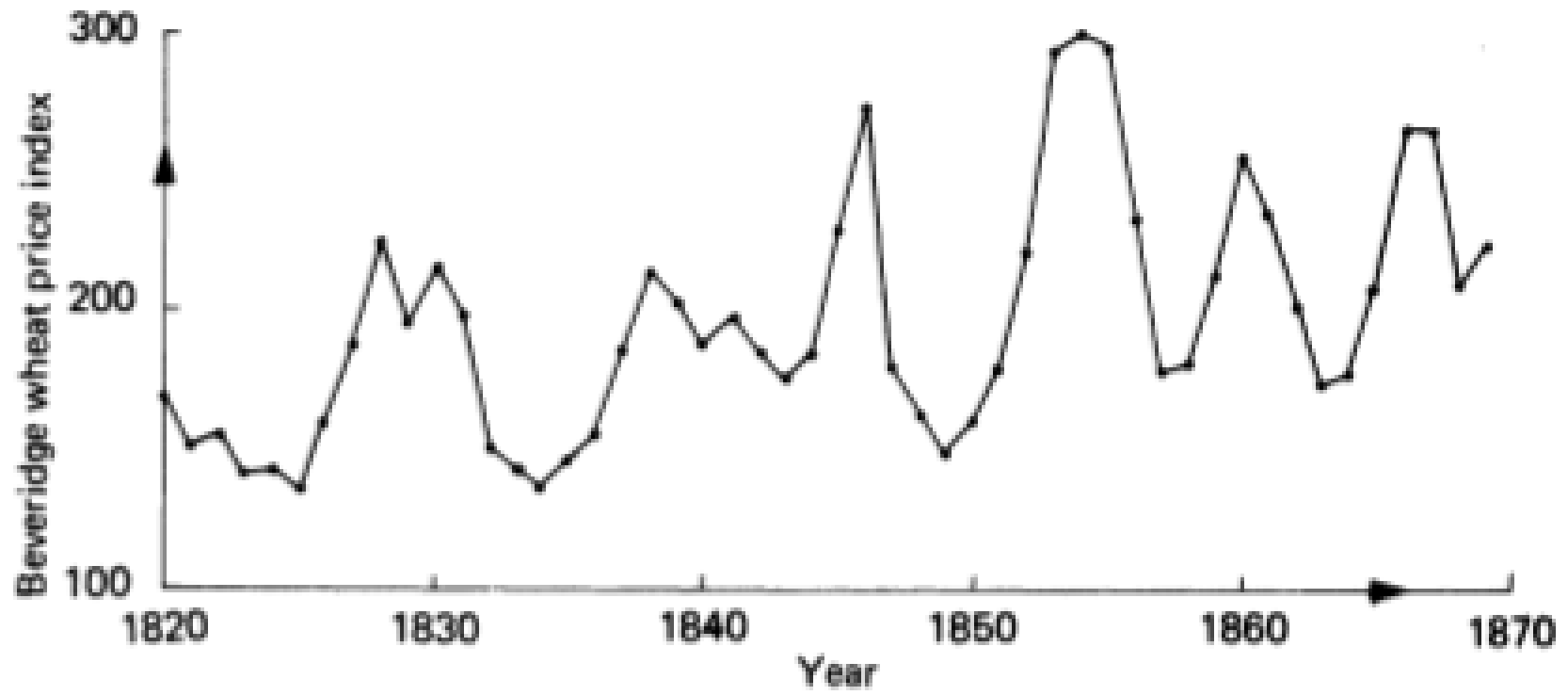
Σειρές δεδομένων που περιγράφουν την δυναμική εξέλιξη ενός φαινομένου.

Οι τιμές μίας παραμέτρου οι οποίες αποθηκεύονται σε ένα διάνυσμα βρίσκονται σε χρονολογική σειρά, συνεπώς η απόλυτη θέση τους έχει μεγάλη σημασία.

Η κάθε τιμή του διανύσματος αντιστοιχεί σε μία καθορισμένη χρονική στιγμή.

Η απόσταση μεταξύ των διαδοχικών χρονικών στιγμών (Δt) είναι πάντα σταθερή.

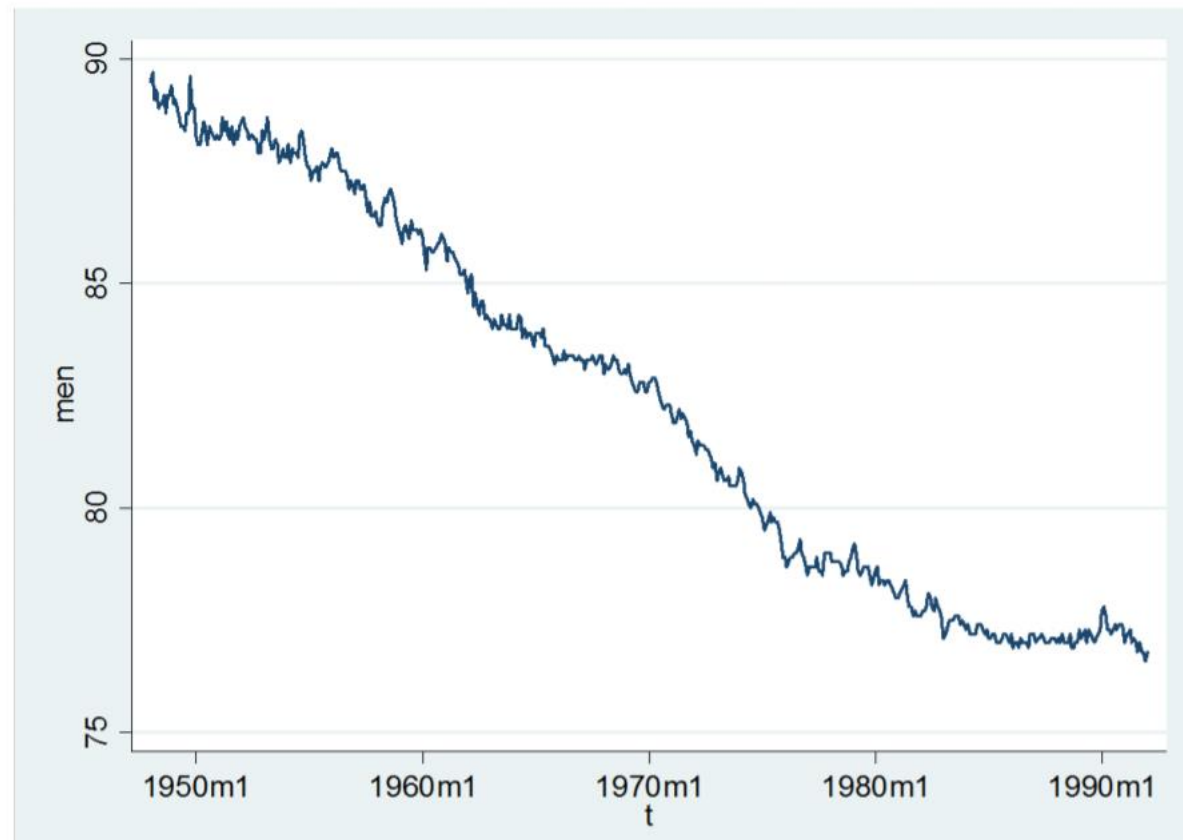
Συνεπώς, η χρονοσειρά είναι ένα σύνολο παρατηρήσεων που ελήφθησαν σε σταθερά χρονικά διαστήματα (Δt , βήμα χρονοσειράς) από μία θέση με στόχο να περιγράψουν την δυναμική εξέλιξη φυσικών φαινομένων.



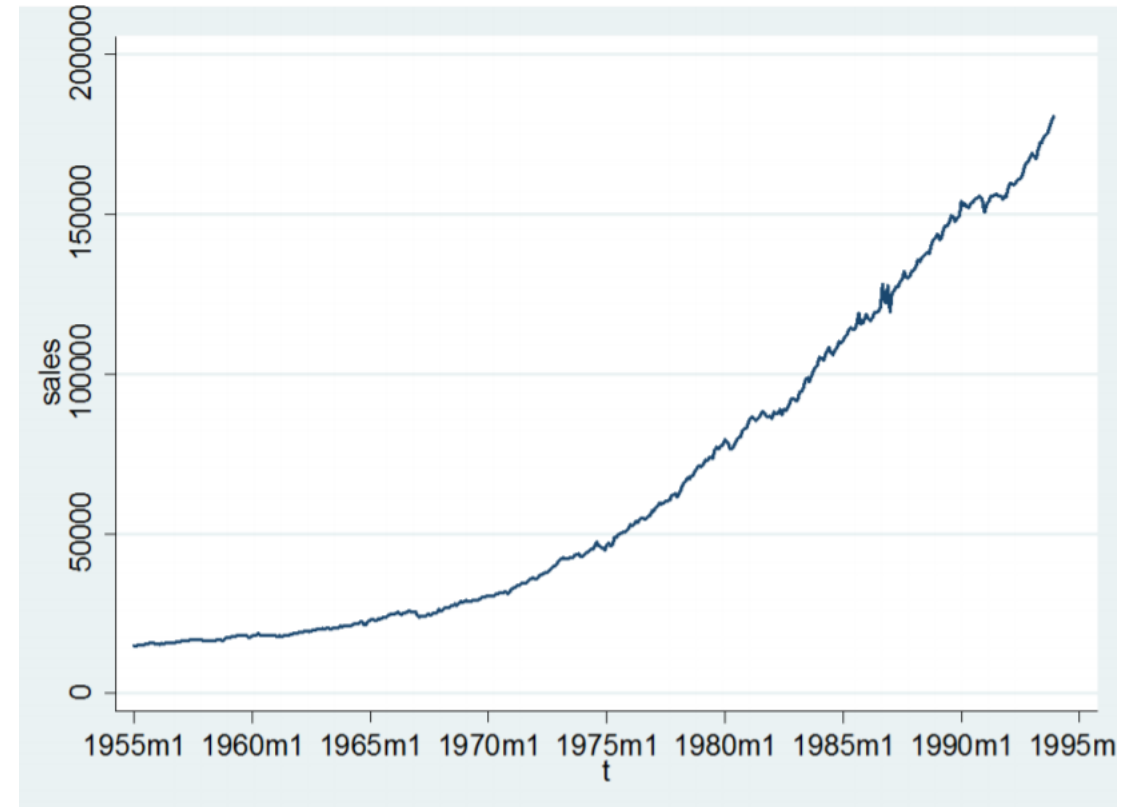
Χαρακτηριστικά μίας Χρονοσειράς

1. Η Τάση (Trend) θα μπορούσε να ορισθεί ως μια μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών μιας χρονοσειράς.

Έτσι, η τάση των τιμών μπορεί να είναι ανοδική, πτωτική ή σταθερή σε ένα συγκεκριμένο χρονικό διάστημα.



Θα μπορούσαμε να προσαρμόσουμε στην τάση της χρονοσειράς μία ή περισσότερες από τις διάφορες οικογένειες καμπυλών, όπως μια ευθεία γραμμή ή μια εκθετική καμπύλη.

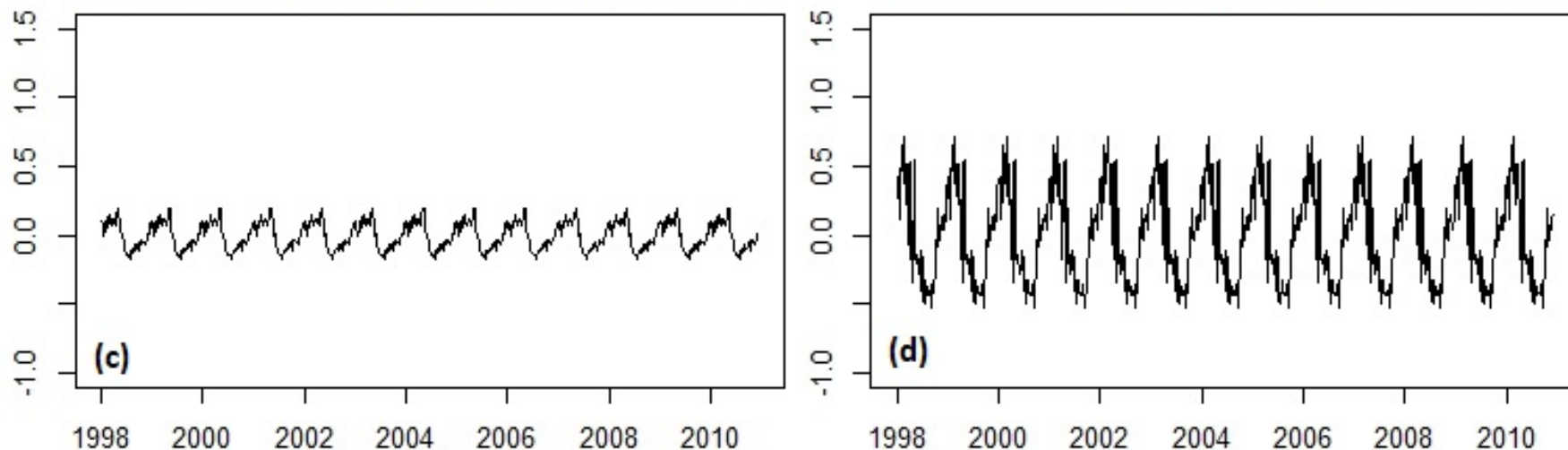


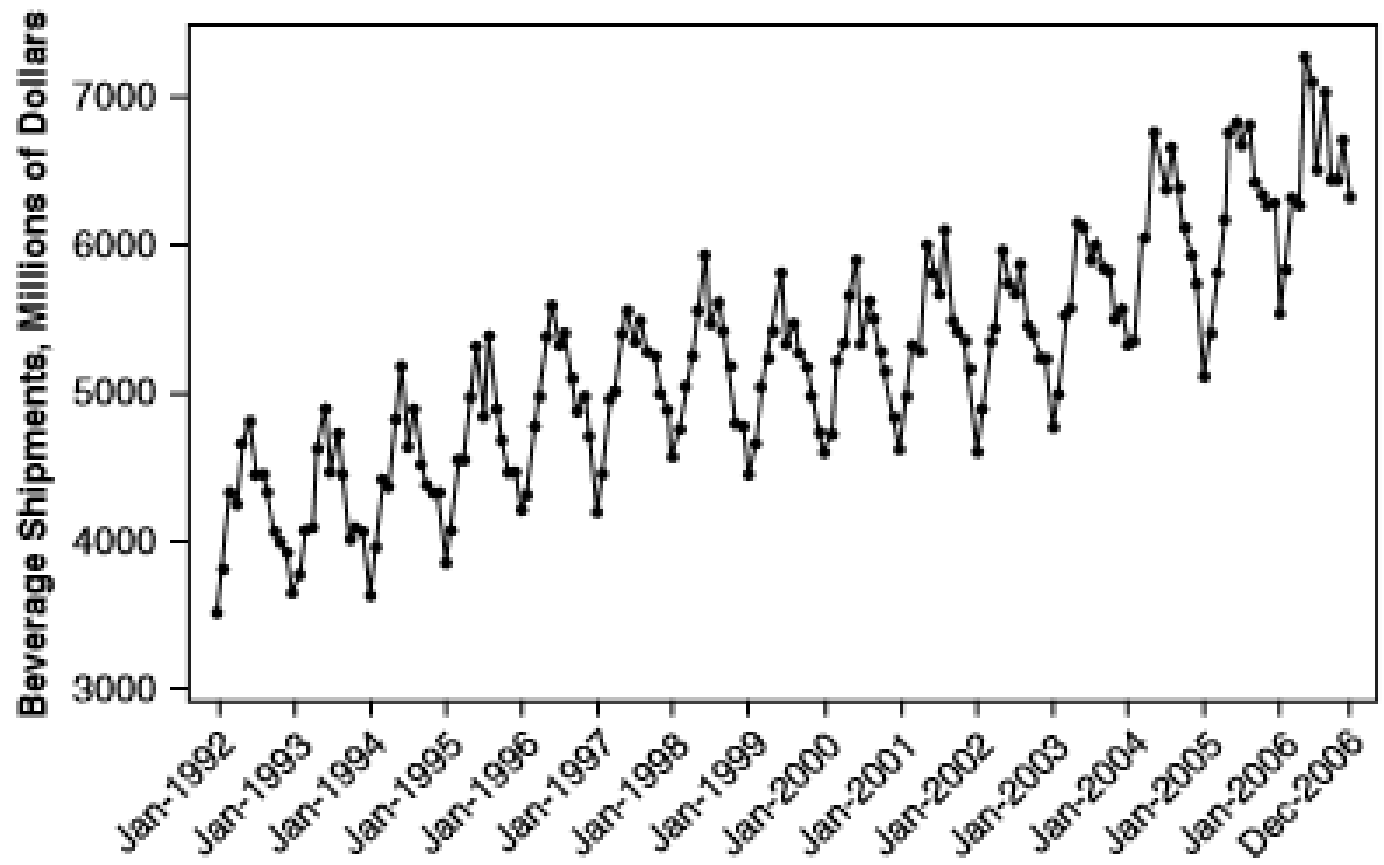
Χαρακτηριστικά μίας Χρονοσειράς

2. Η **Κυκλικότητα ή Εποχικότητα (Seasonality)** αντιπροσωπεύει μια μεταβολή που εμφανίζεται λόγω εξωγενών παραγόντων εμφανίζει περιοδικότητα.

Στις περιβαλλοντικές χρονοσειρές, οι περίοδοι αυτοί είναι συνήθως ετήσιες. Επομένως, η κυκλικότητα/περιοδικότητα οφείλεται στην αλλαγή των εποχών (εξωγενής παράγοντας), οπότε η περίοδος της μεταβολής είναι ένα έτος.

Στις γραφικές παραστάσεις των χρονοσειρών παρουσιάζεται ως μια κυματοειδής γραμμή που κινείται ανάμεσα στην υψηλότερη και χαμηλότερη στάθμη.



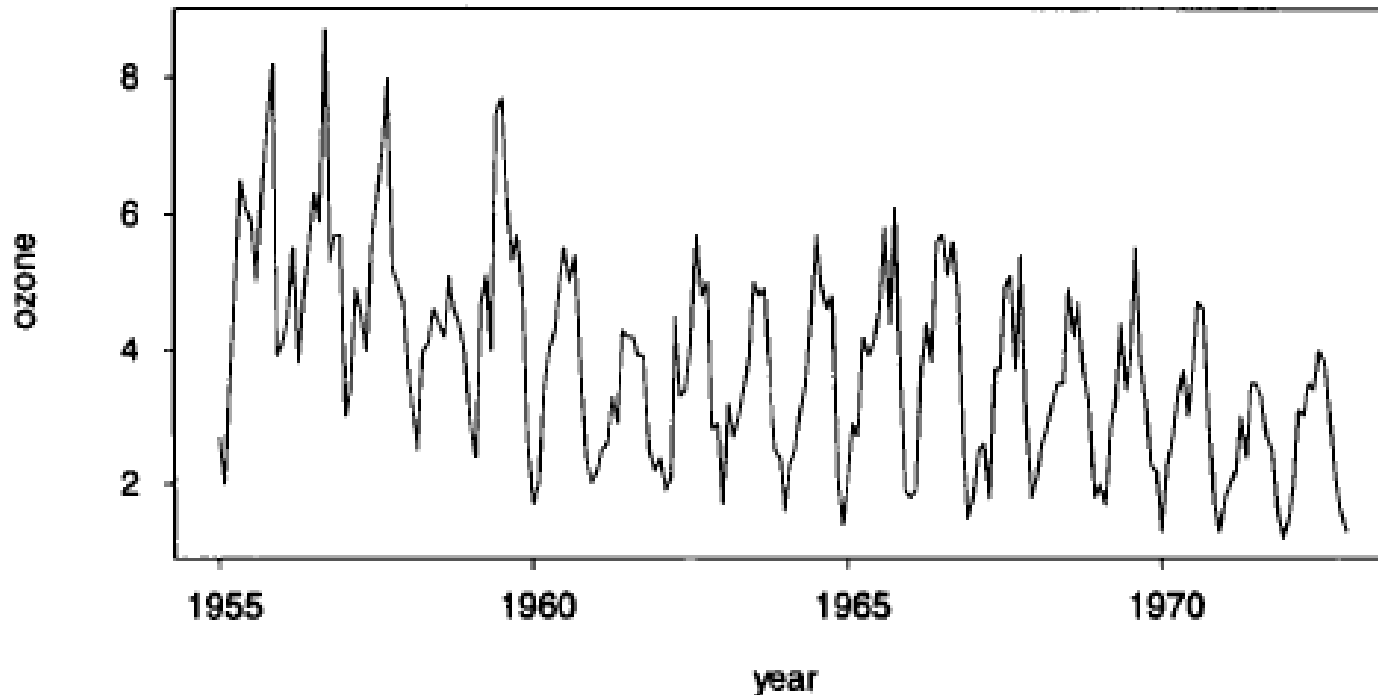


Χρονοσειρά με εποχικότητα και τάση.

Η διακύμανση αυτή είναι άμεσα κατανοητή και προβλέψιμη, διότι τα δεδομένα ορισμένων χρονοσειρών επαναλαμβάνονται με τον ίδιο περίπου τρόπο σε σχέση με το χρόνο.

Εφόσον, η εποχική διακύμανση παρουσιάζεται με συστηματικό τρόπο, είναι ένα χαρακτηριστικό εύκολα οπτικά αναγνωρίσιμο που μπορεί να μετρηθεί και να απομονωθεί, ώστε να μην επηρεάζει τα δεδομένα μας.

Η νέα χρονοσειρά που προκύπτει ονομάζεται **από-εποχικοποιημένη χρονοσειρά**.



Μέσες μηνιαίες τιμές του όζοντος στο κέντρο του Λος Άντζελες από το 1955 μέχρι το 1972.

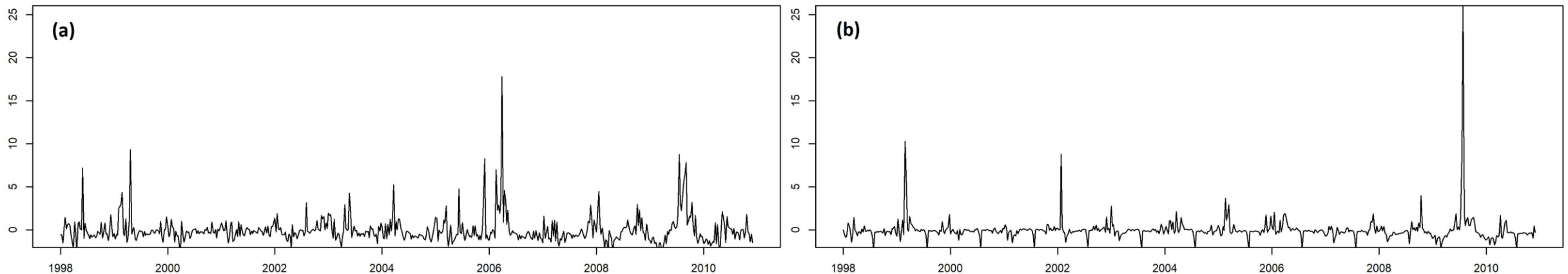
Παρατηρείται ότι, το ατμοσφαιρικό όζον που είναι ένας δείκτης της ατμοσφαιρικής ρύπανσης παρουσιάζει έντονη εποχικότητα, η οποία είναι υψηλή κατά τους καλοκαιρινούς μήνες και χαμηλή τον χειμώνα.

Χαρακτηριστικά μίας Χρονοσειράς

3. **Υπολοιπόμενες Ακραίες τιμές (Outliers)** είναι οι τιμές που προκύπτουν αν από την χρονοσειρά αφαιρέσουμε την Τάση και την Εποχικότητα.

Είναι απομονωμένες παρατηρήσεις που εμφανίζονται στο γράφημα κάποιας χρονοσειράς ως απότομες αλλαγές στο πρότυπο συμπεριφοράς της.

Οι ακραίες τιμές είναι μη προβλέψιμες και η επίδρασή τους στην χρονοσειρά έχει μικρή χρονική διάρκεια. Ένα outlier μπορεί να αντιπροσωπεύει μια ασυνήθιστη παρατήρηση που οφείλεται σε κάποιο απρόβλεπτο γεγονός.



Χαρακτηριστικά μίας Χρονοσειράς

Άρα, κάθε χρονοσειρά αποτελείται από την Τάση (T), την Εποχικότητα (S) και τις Ακραίες Τιμές (R).

$$Y_t = T_t + S_t + R_t$$

Βήμα 1. Θα ανοίξουμε το e-class και θα κατεβάσουμε σε ένα κατάλογο εργασίας το αρχείο temperature.xlsx το οποίο περιέχει δυο στήλες:

- Η πρώτη στήλη τις ημερομηνίες
- Η δεύτερη στήλη την θερμοκρασία της θάλασσας

Βήμα 2. Θα δημιουργήσουμε στην επιφάνεια εργασίας έναν κατάλογο με όνομα "test". Θα αποθηκεύσουμε εκεί το αρχείο temperature.xlsx.

Βήμα 3. Θα κάνουμε Αντιγραφή και τις δύο στήλες από το Excel και θα τις Επικολλήσουμε στο Σημειωματάριο (αρχείο temperature.txt).

Βήμα 4. Θα αποθηκεύσουμε το αρχείο temperature.xlsx ως temperature.csv

Όλα τα αρχεία θα πρέπει να βρίσκονται στον ίδιο κατάλογο test.

Βήμα 5. Στην R ανοίγω το script και το σώζω στον ίδιο κατάλογο test.

Βήμα 6. Δηλώνω τον κατάλογο εργασίας

```
# Find the working directory  
setwd("C:/Users/user/Desktop/test")
```

Βήμα 7. Διαβάζουμε τα δεδομένα από το αρχείο κειμένου

```
#read the *.txt file  
data = read.table("temperature.txt", header=TRUE)
```

```
temp = data$X.C.
```

Βήμα 8. Διαβάζουμε τα δεδομένα από το αρχείο csv

```
#read the *.csv file  
data1 = read.csv("temperature.csv", header = TRUE, sep=";")
```

```
temper = data1$X.C.
```

Βήμα 9. Δημιουργώ την χρονοσειρά

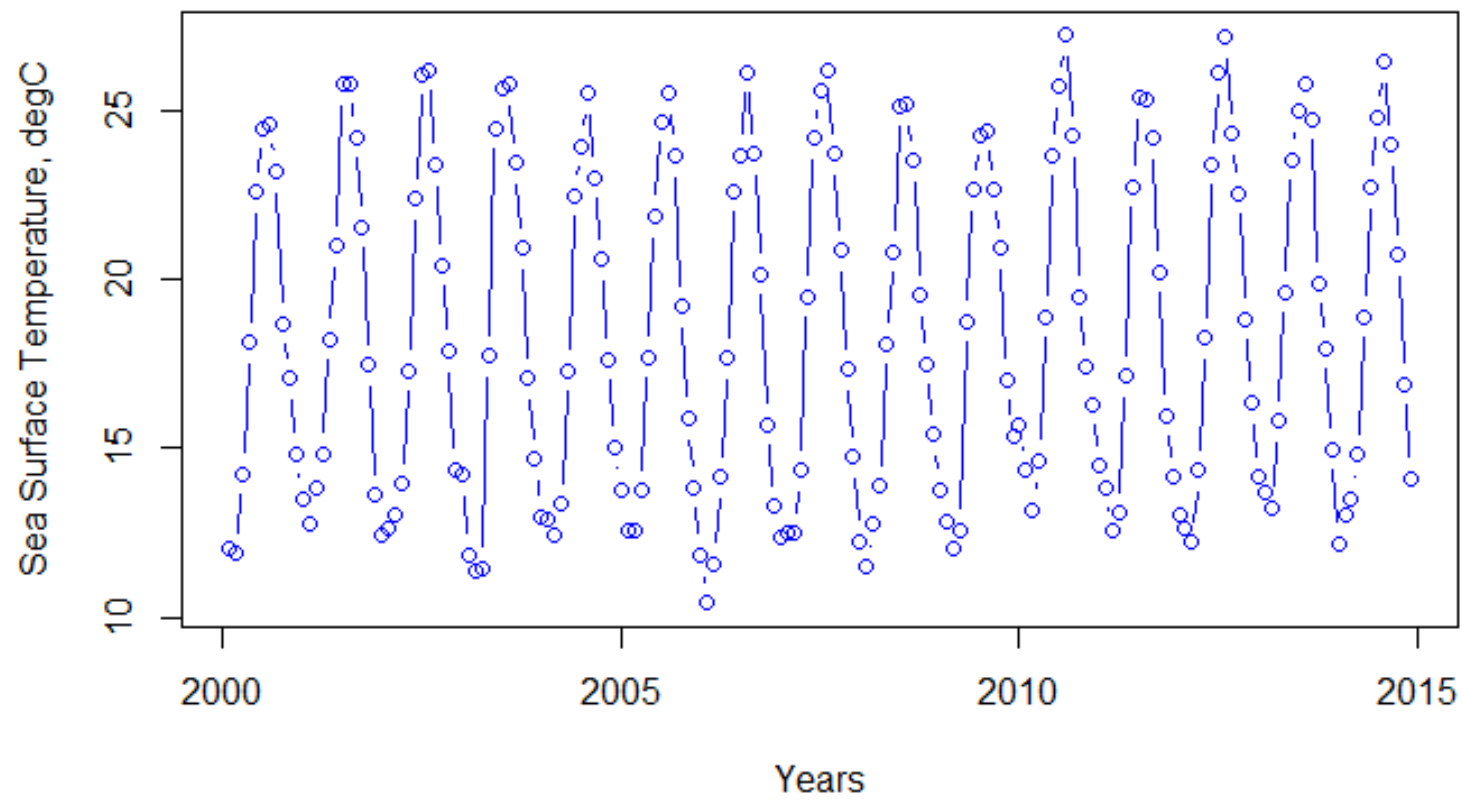
```
temper.ts = ts(temper, frequency = 12, start = c(2000,2))
```

Βήμα 10. Δημιουργώ γράφημα της χρονοσειράς

```
plot(temper.ts,  
     xlab = "Years",  
     ylab = "Sea Surface Temperature, degC",  
     lty=3,  
     col=4,  
     type="b")
```

Βήμα 11. Κάνω στατιστική ανάλυση στην χρονοσειρά
summary(temper.ts)

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|-------|
| 10.41 | 13.83 | 17.67 | 18.39 | 23.42 | 27.22 |



Βήμα 12. Για να πάρουμε τα χαρακτηριστικά της χρονοσειράς γράφουμε:

```
attributes(temper.ts)
```

```
$tsp
```

```
[1] 2000.083 2014.917 12.000
```

```
$class
```

```
[1] "ts"
```

που μας επιστρέφει ότι η χρονοσειρά ξεκινά στην αρχή του 2000 και λήγει στο τέλος του 2014.

Ο αριθμός 12 σημαίνει ότι υπάρχουν 12 τιμές σε κάθε χρόνο, άρα η χρονοσειρά έχει μηνιαίες τιμές.

Εφαρμογή Φίλτρων σε Χρονοσειρά

Μάθαμε ότι το κύριο στοιχείο ανάλυσης μίας χρονοσειράς είναι ο διαχωρισμός της χρονοσειράς X_t σε τρία επιμέρους τμήματα:

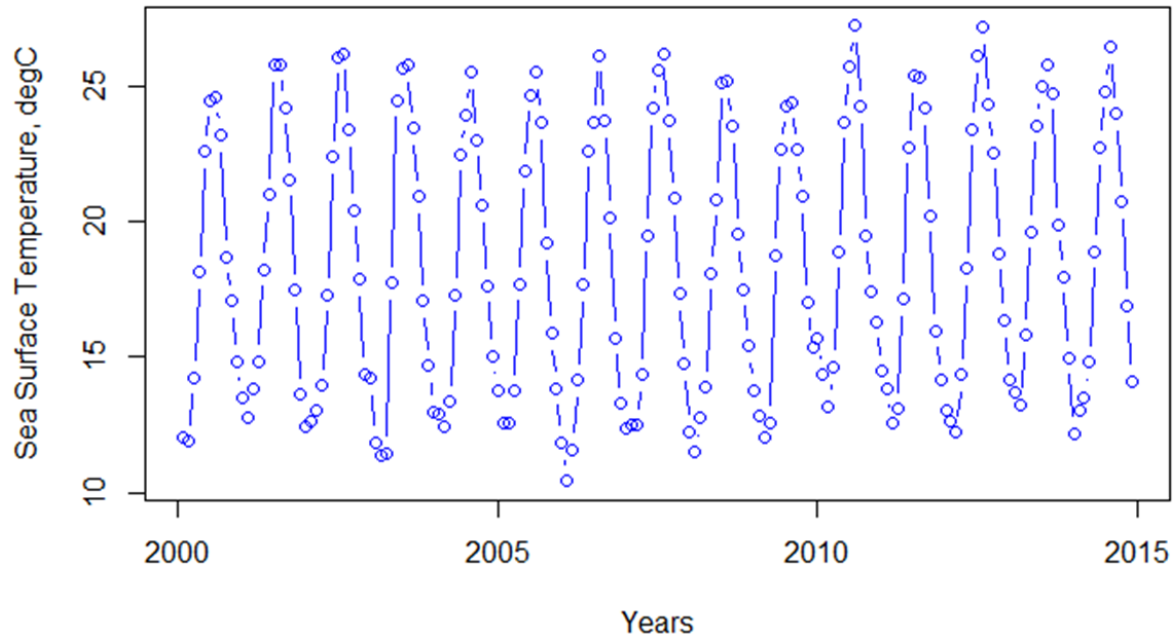
- α) την τάση (T_t),
- β) την εποχικότητα (S_t), και
- γ) τα κατάλοιπα (R_t).

Η κοινή μέθοδος προσδιορισμού της τάσης είναι μέσω της χρήσης κοινών φίλτρων για ένα συγκεκριμένο τμήμα της χρονοσειράς:

$$T_t = \sum_{i=-\infty}^{+\infty} \lambda_i X_{t+i}$$

Μία απλή ομάδα γραμμικών φίλτρων είναι οι κινούμενοι μέσοι (moving averages) με ίσα βάρη:

$$T_t = \frac{1}{2a+1} \sum_{i=-a}^{+a} X_{t+i}$$



$$T_t = \frac{1}{2a+1} \sum_{i=-a}^{+a} X_{t+i}$$

Στη περίπτωση αυτή, οι φιλτραρισμένες τιμές της χρονοσειράς για μια συγκεκριμένη περίοδο προκύπτουν από τη μέση τιμή των τιμών

$$\{x_{t-a}, \dots, x_t, \dots, x_{t+a}\}.$$

Οι συντελεστές εφαρμογής του φίλτρου είναι $\{1/(2a+1), \dots, 1/(2a+1)\}$

$$a=2, \lambda_i = \left[\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5} \right]$$

$$a=8, \lambda_i = \left[\frac{1}{17}, \frac{1}{17}, \dots, \frac{1}{17}, \frac{1}{17}, \frac{1}{17} \right], 17 \text{ times}$$

$$a=12, \lambda_i = \left[\frac{1}{25}, \frac{1}{25}, \dots, \frac{1}{25}, \frac{1}{25} \right], 25 \text{ times}$$

Η εντολή για την εφαρμογή του φίλτρου στη χρονοσειρά είναι `filter()`.

```
#Time-series filtering
```

```
# alfa = 2
```

```
new1 = filter(temper.ts, filter=rep(1/5,5))
```

```
lines(new1,col='red')
```

```
#alfa = 8
```

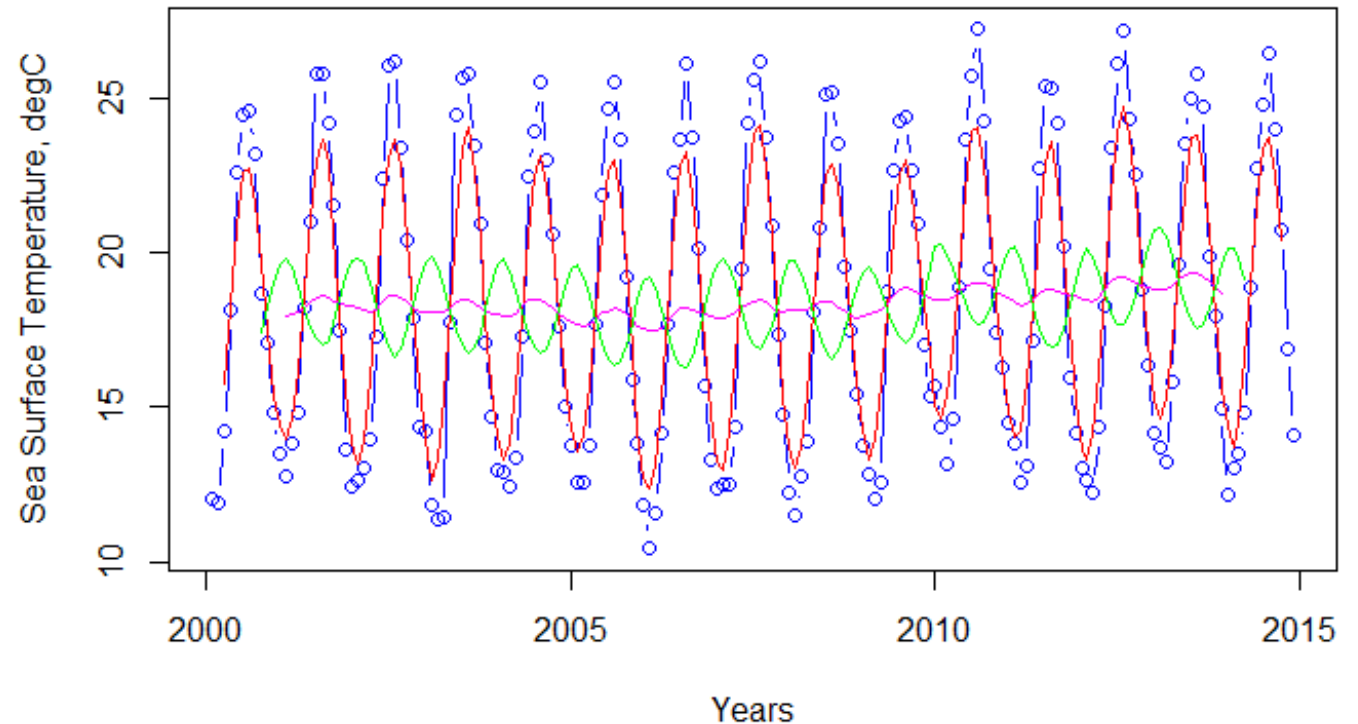
```
new2 = filter(temper.ts, filter=rep(1/17,17))
```

```
lines(new2,col='green')
```

```
#alfa = 12
```

```
new3 = filter(temper.ts, filter=rep(1/25,25))
```

```
lines(new3,col='magenta')
```



Διαχωρισμός χρονοσειράς σε επιμέρους τμήματα – Τάση, Εποχικότητα, Κατάλοιπα

Για να διαχωρίσω την χρονοσειρά `temper.ts` στα επιμέρους τμήματα που την συνθέτουν θα χρησιμοποιήσω την συνάρτηση `stl`

```
stl(temper.ts, s.window="per")
```

Μπορώ να αποδώσω τις επιμέρους τρεις χρονοσειρές σε ένα νέο αντικείμενο

```
new = stl(temper.ts, s.window="per")
```

Αν γράψω `new` στην γραμμή εντολής παίρνω τις τιμές των τριών χρονοσειρών.

| | seasonal | trend | remainder |
|----------|------------|----------|--------------|
| Feb 2000 | -5.7085838 | 17.80595 | -0.042368975 |
| Mar 2000 | -5.7696628 | 17.82966 | -0.166994081 |
| Apr 2000 | -4.4181420 | 17.85336 | 0.803781000 |
| May 2000 | -0.1455158 | 17.88911 | 0.385410594 |
| Jun 2000 | 4.3781102 | 17.92485 | 0.261040392 |
| Jul 2000 | 6.7183016 | 17.97141 | -0.218712508 |
| Aug 2000 | 7.4394267 | 18.01797 | -0.857399085 |
| Sep 2000 | 5.3579252 | 18.06237 | -0.228294217 |

Διαχωρισμός χρονοσειράς σε επιμέρους τμήματα – Τάση, Εποχικότητα, Κατάλοιπα

Παρατηρώ το αντικείμενο new. Είναι τύπου list.

Αποτελείται από επιμέρους αντικείμενα, όπως time.series, weights, call, win, deg, κλπ.

Για να καλέσω την χρονοσειρά της εποχικότητας θα γράψω

```
new$time.series[, 1]
```

Για να καλέσω την χρονοσειρά της τάσης θα γράψω

```
new$time.series[, 2]
```

Για να καλέσω την χρονοσειρά των καταλοίπων θα γράψω

```
new$time.series[, 3]
```

Διαχωρισμός χρονοσειράς σε επιμέρους τμήματα – Τάση, Εποχικότητα, Κατάλοιπα

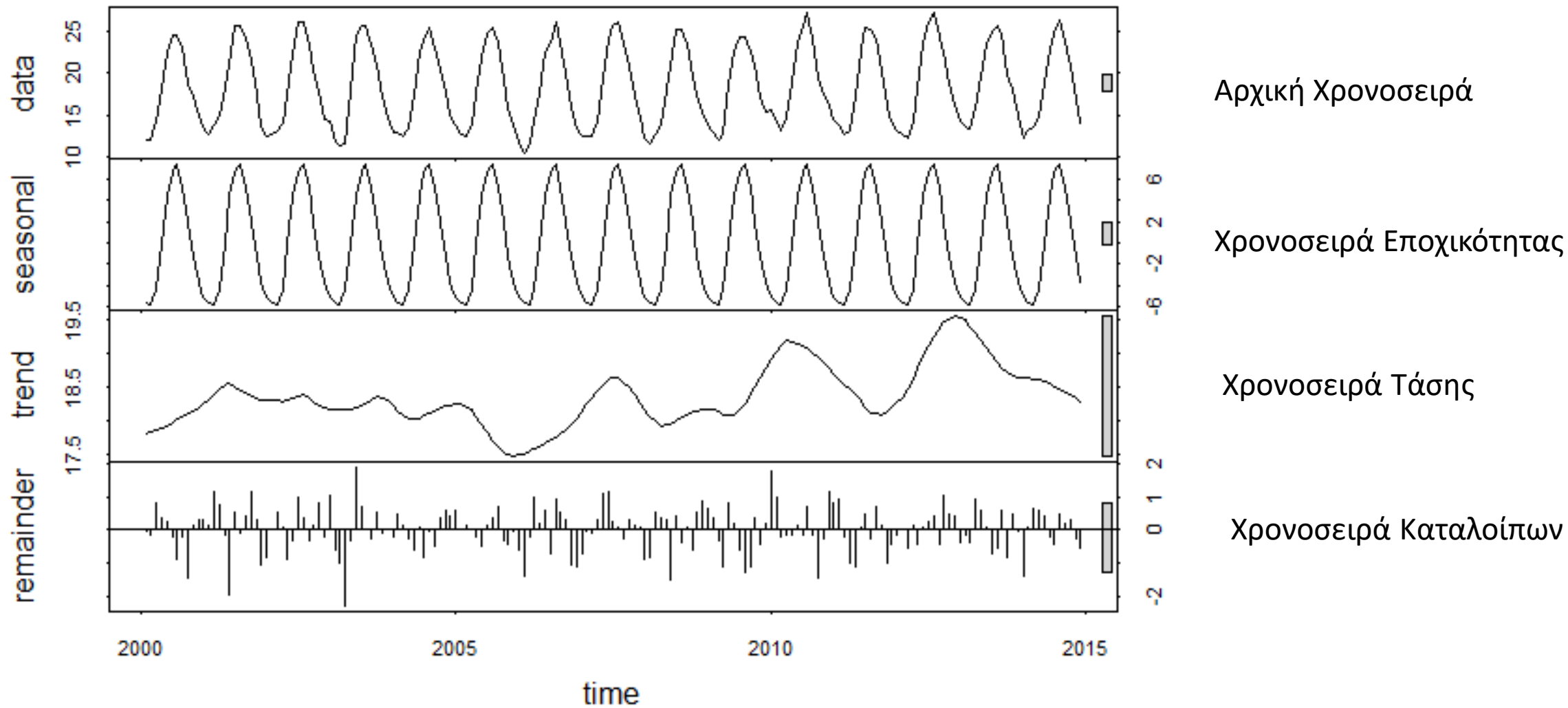
Για να κάνω το διάγραμμα της χρονοσειράς της τάσης

```
plot(new$time.series[,2])
```

Με τον ίδιο τρόπο και για τις άλλες χρονοσειρές.

Για να κάνω ένα συγκεντρωτικό διάγραμμα όπου να βρίσκονται όλες οι χρονοσειρές και η αρχική χρονοσειρά γράφω:

```
plot(new)
```



Εφαρμογή Γραμμικού Μοντέλου για τον Προσδιορισμό της Τάσης

Θα δημιουργήσουμε ένα διάνυσμα που θα μετρά τις τιμές της χρονοσειράς

```
t = seq(1,length(temper.ts))
```

Θα εφαρμόσουμε ένα γραμμικό μοντέλο επί της χρονοσειράς θερμοκρασιών

```
linear = lm(temper.ts~t)
```

```
plot(t,temp,type="l")
```

```
lines(lm(temper.ts~t)$fit,col=2)
```

Το αντικείμενο `linear` είναι λίστα που περιέχει τους συντελεστές της γραμμικής παλινδρόμησης.

```
linear$coefficients
```

