

Trabajo Final - Series de Tiempo

Paula Beati, Damian Farias, Juan Pablo Bresciani, Nestor Coppolillo

Julio 2016

Universidad de Buenos Aires

Maestria en Explotacion de Datos y Descubrimiento de Conocimiento

Profesor: Marcelo Risk

Resumen

El trabajo final de la materia Series de Tiempo consiste en aplicar las tecnicas aprendidas sobre un dataset, observar los resultados y obtener conclusiones.

En nuestro caso, seleccionamos series de exportaciones argentinas durante los años 2014 y 2015. En principio realizamos un analisis exploratorio de los datos comparando distintos intervalos de tiempo: diario, semanal y mensual.

Algunas series se transformaron al dominio frecuencia, se aplicaron distintos filtros y fueron vueltos a generar en el dominio tiempo con la transformada inversa para compararlas con su formato original.

Se realizaron correlaciones entre series de exportaciones de un mismo producto con igual rango e intervalo de tiempo y distintos lugares de salida, buscando predicciones entre ellas.

Introducción

Las exportaciones argentinas son datos que se pueden presentar como series de tiempo, considerando el importe, la cantidad y el peso de cada operacion de comercio exterior, agrupados por dia. Estas tres magnitudes permiten un analisis de la informacion desde distintos puntos de vista.

Como las exportaciones se expresan en un nivel muy detallado, para facilitar la comprensión de los resultados las agrupamos por dos atributos: el lugar de origen (la aduana de salida del producto) y el tipo de mercadería a nivel de capitulo, que los clasifica en 99 tipos distintos.

Hipótesis y Objetivos

Material y Métodos

Los datos son series con la siguiente estructura:

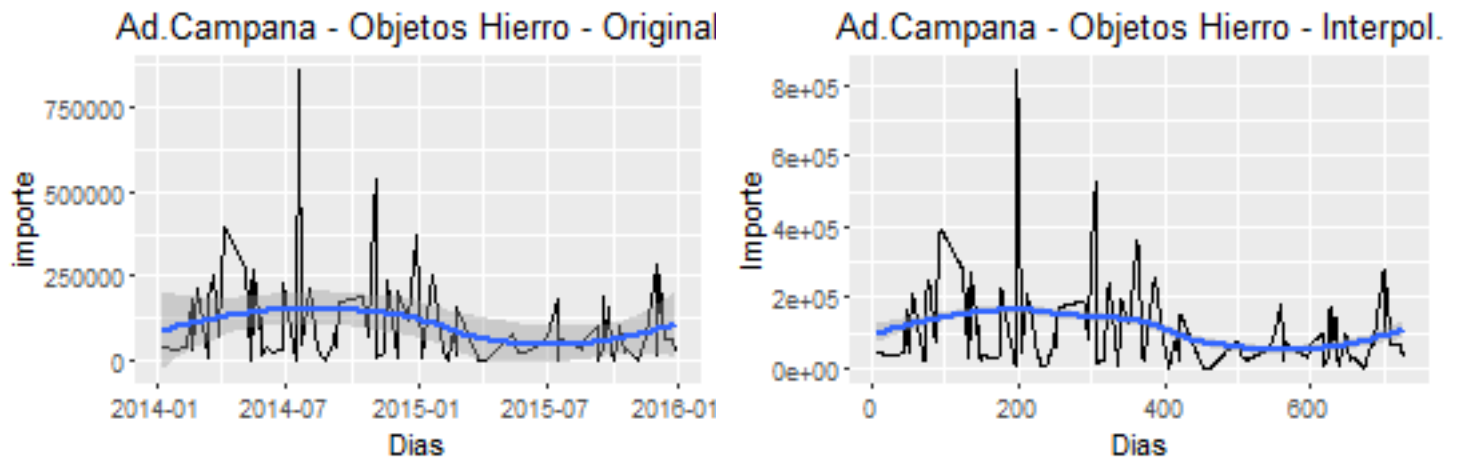
- Aduana. (por ej: Mar del Plata, Buenos Aires, etc.)
- Tipo de mercaderia (carne bovina, manzanas, minerales, pinturas, etc.)
- Fecha de exportacion (por dia)
- Cantidad
- Kilos
- Importe

Para procesar estos datos generamos funciones usando paquetes y el lenguaje R.

Pre-procesamiento de los datos

Como los datos fueron sumariados por aduana-mercaderia-dia para todas las exportaciones, se filtraron las series que tuvieran pocos datos. Por otro lado, dado que algunas herramientas de analisis presuponen que las muestras fueron tomadas en forma equiespaciada en el tiempo, se rellenaron los dias donde no hubo registro de exportacion usando interpolacion.

Ejemplo de serie con los datos originales y luego de la interpolacion



Resultados

Generacion de series usando distintos intervalos de tiempo

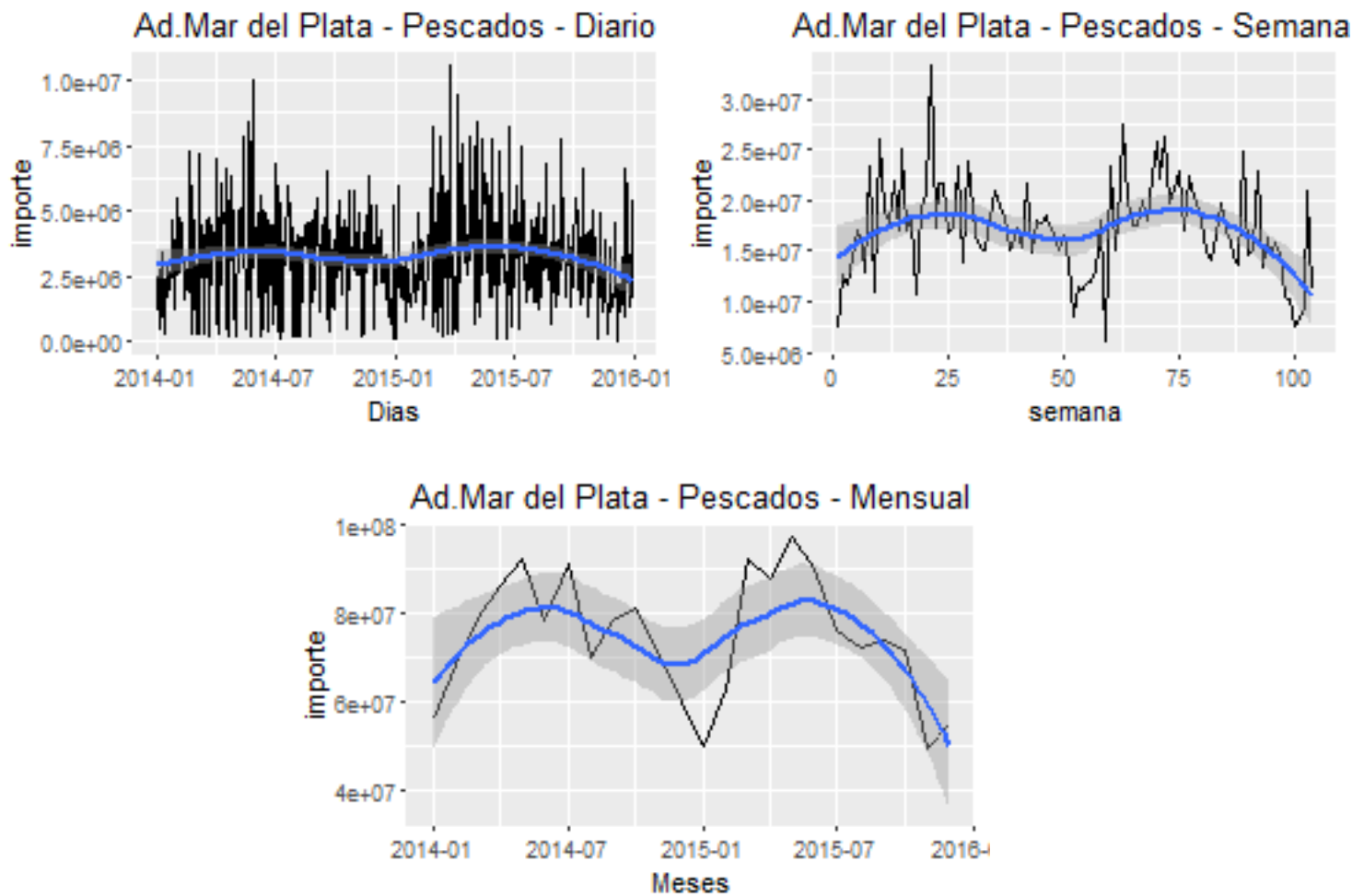


Figure 1: Intervalo mensual

Aplicacion de filtros a la series

Deteccion y filtrado de estacionalidades

Consideramos una serie diaria de exportaciones de pescado en la aduana Mar del Plata 37 para los años 2014 y 2015. Primero se muestra con los datos originales, donde no tenemos datos todos los dias; el plot parece rellenar los dias faltantes. El proximo grafico muestra la serie equiespaciada, “con relleno”, aunque sin dato de importe. El otro si tienen los dias faltantes un valor interpolado linealmente.

La serie interpolada es pasada al dominio de la frecuencia y se indica el pico maximo y la frecuencia en el mismo, cada 7 dias. Le aplicamos un filtro pasa bajos para los valores que estan afuera del intervalo entre los dos picos maximos y recontruimos la serie en el dominio del tiempo, sin estacionalidad.

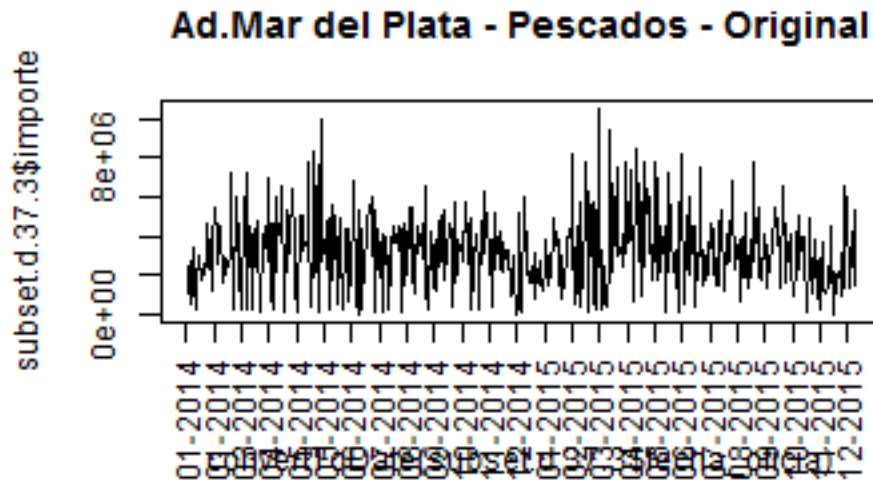


Figure 2: Filtrado de estacionalidades

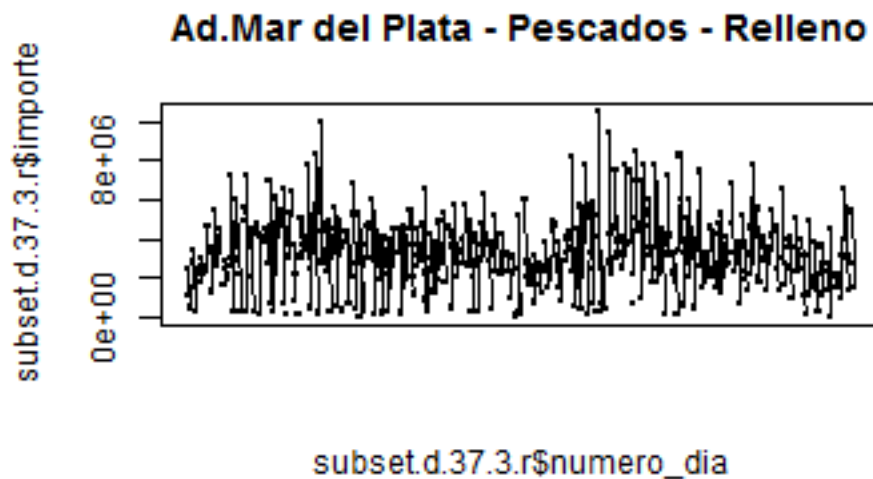


Figure 3: Filtrado de estacionalidades

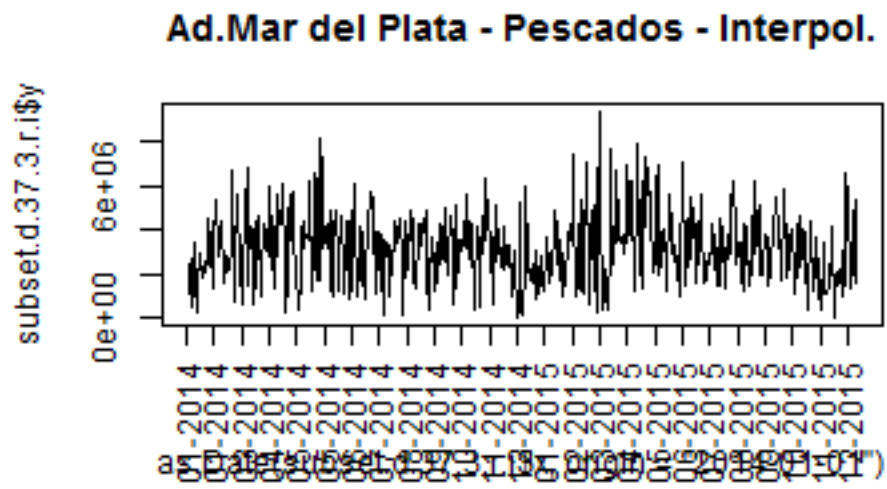


Figure 4: Filtrado de estacionalidades

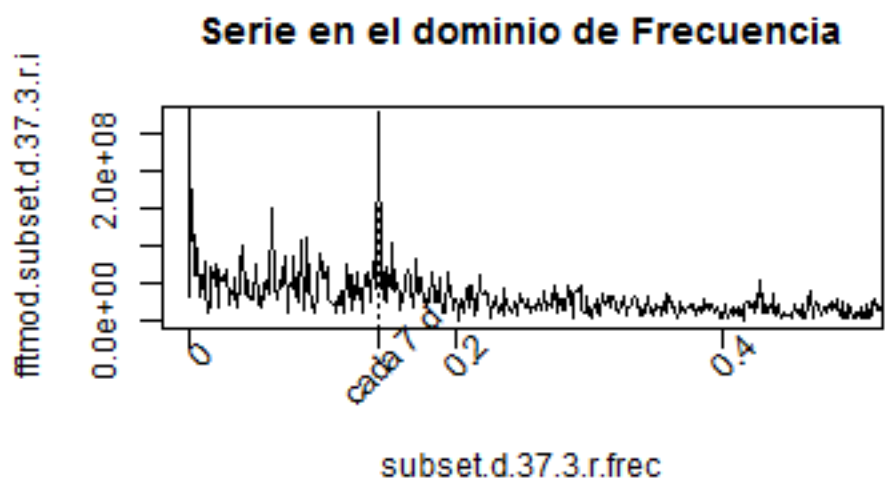


Figure 5: Filtrado de estacionalidades

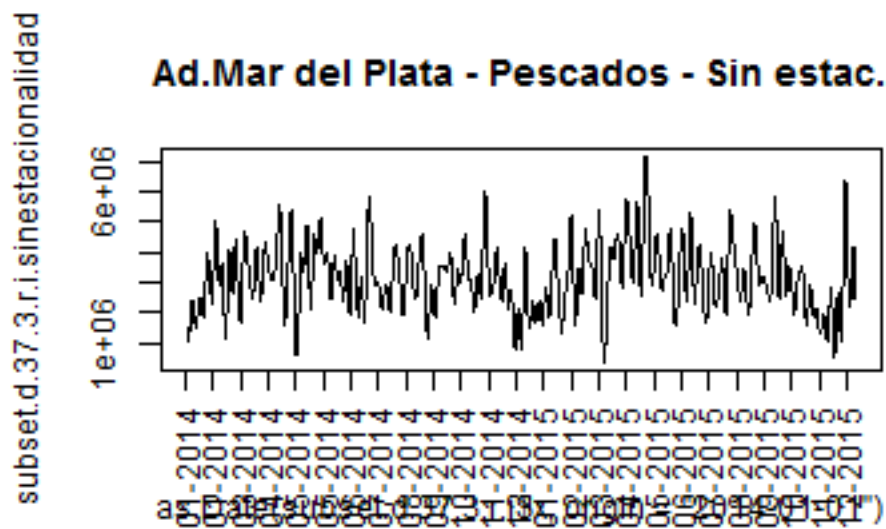
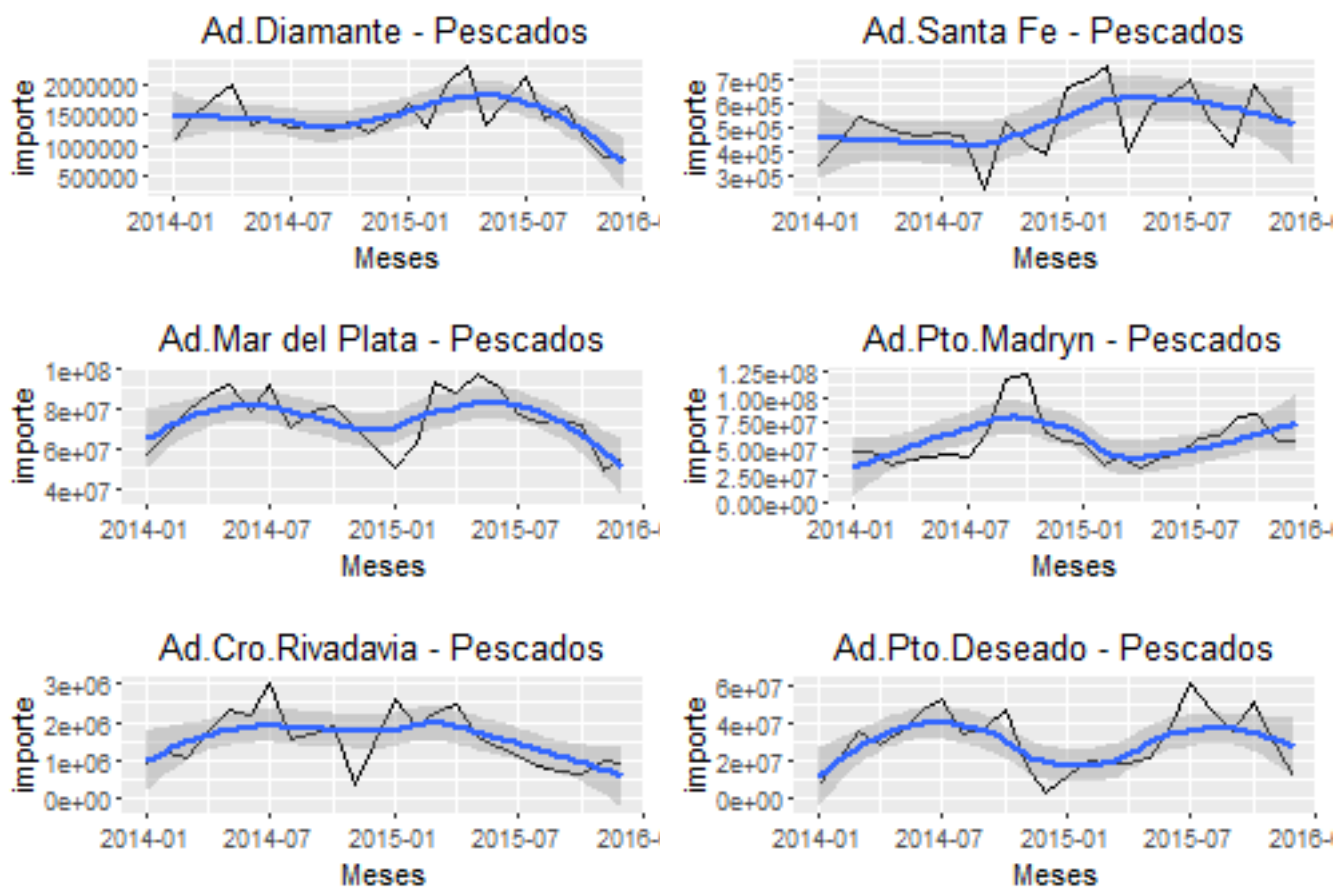
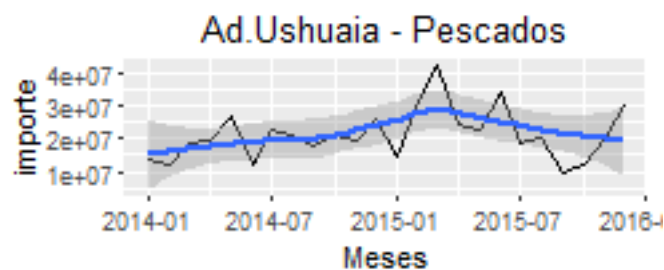
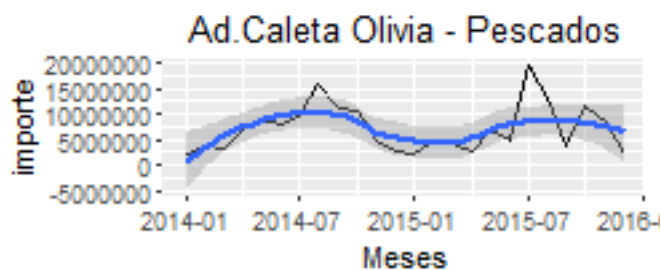


Figure 6: Filtrado de estacionalidades

Correlacion

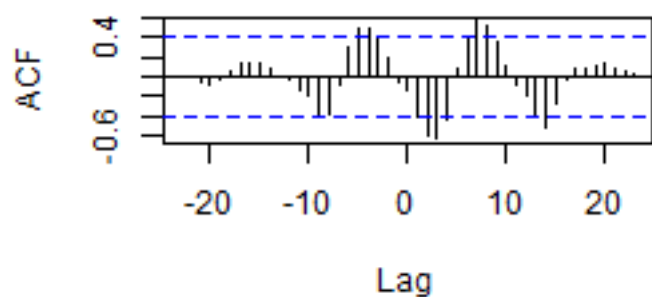
Se consideran series de exportaciones de pescados realizadas en ocho aduanas, seis maritimas y las otras dos cercanas a un río, durante los años 2014 y 2015. Primero vemos las series de las aduanas mensuales por importe.



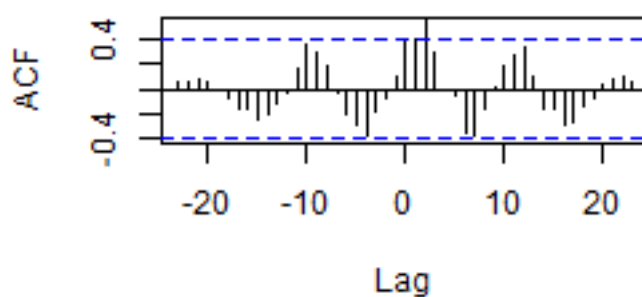


Ahora hacemos una correlacion entre la aduana Mar del Plata y las otras cinco aduanas maritimas.

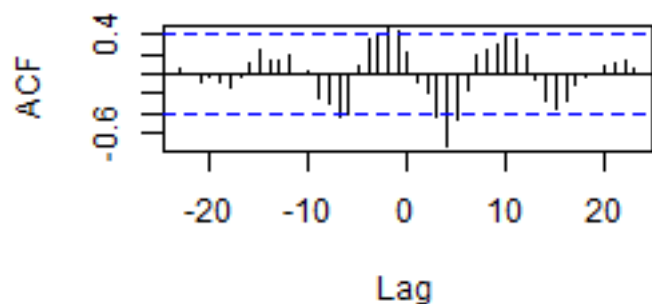
Ad.Mar del Plata - Ad.Pto.Madryn



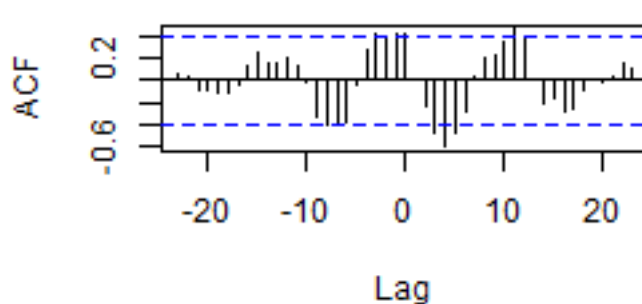
Ad.Mar del Plata - Ad.Cro.Rivadavia



Ad.Mar del Plata - Ad.Caleta Olivia



Ad.Mar del Plata - Ad.Pto.Deseado



Ad.Mar del Plata - Ad.Ushuaia

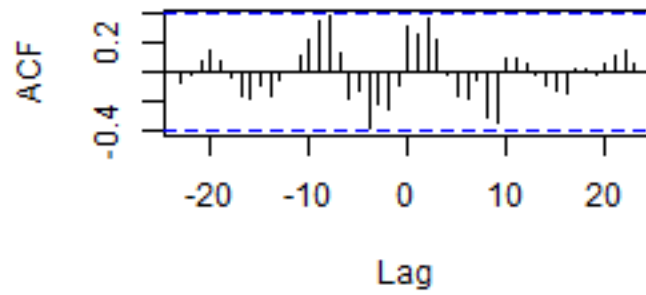
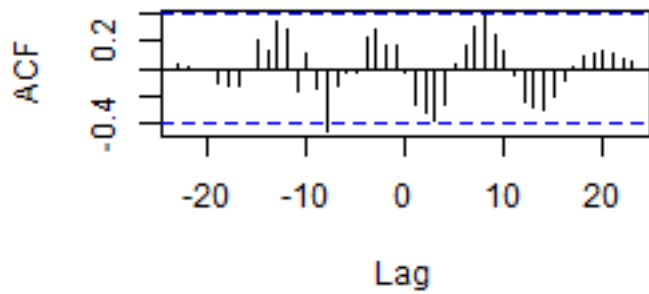


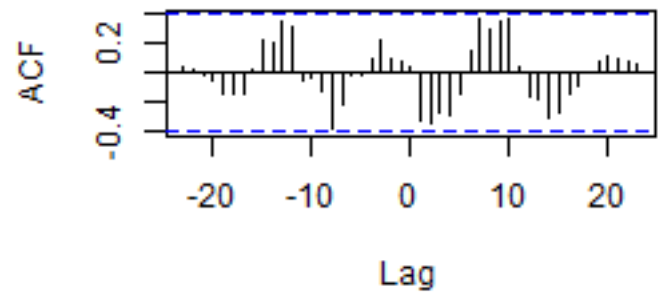
Figure 7: Series mensuales

Y correlacionamos tres ciudades cercanas entre ellas

Ad.Cro.Rivadavia - Ad.Caleta Olivia



Ad.Cro.Rivadavia - Ad.Pto.Deseado



Ad.Caleta Olivia - Ad.Pto.Deseado

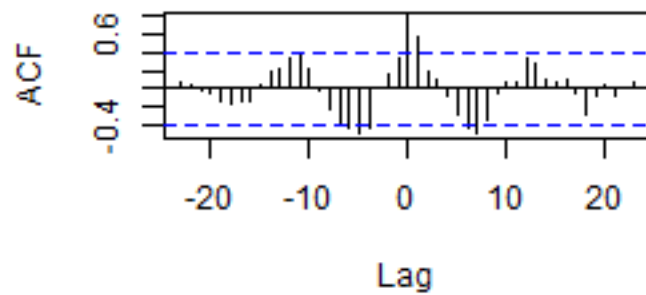


Figure 8: Series mensuales

Y dos ciudades a orillas del rio Parana.

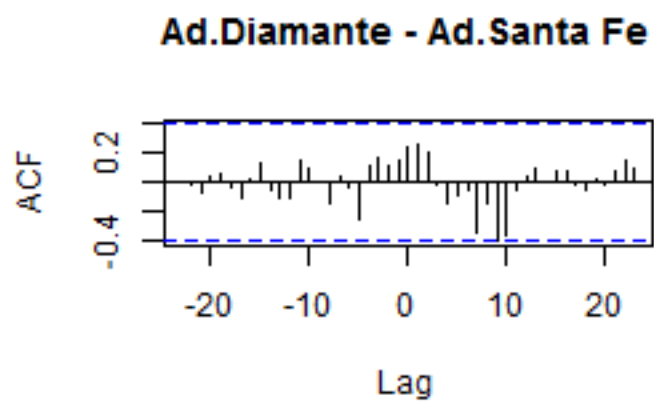
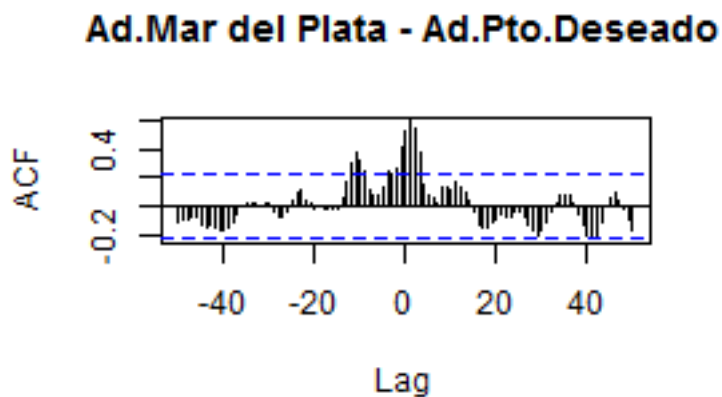
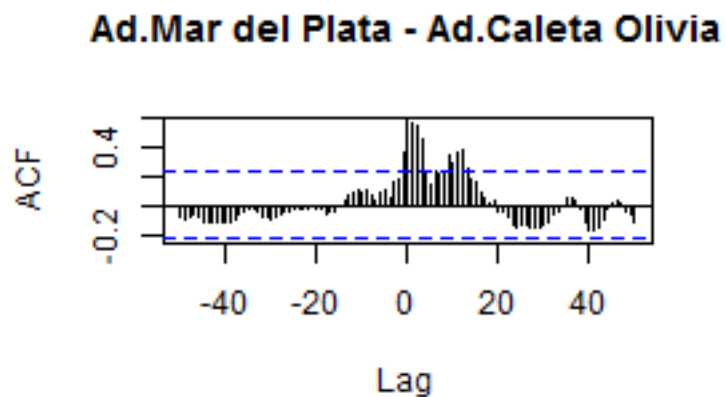
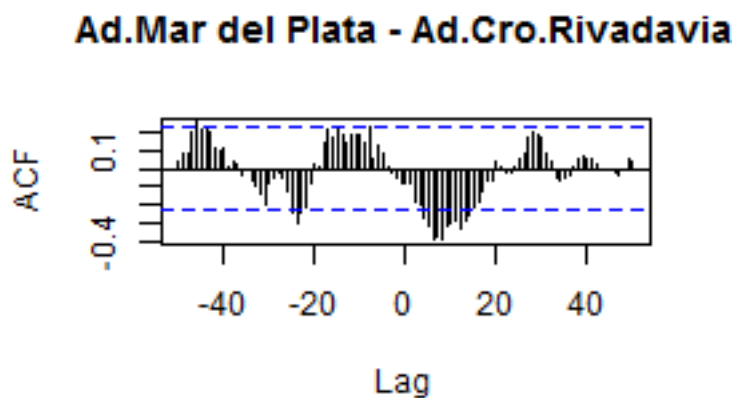
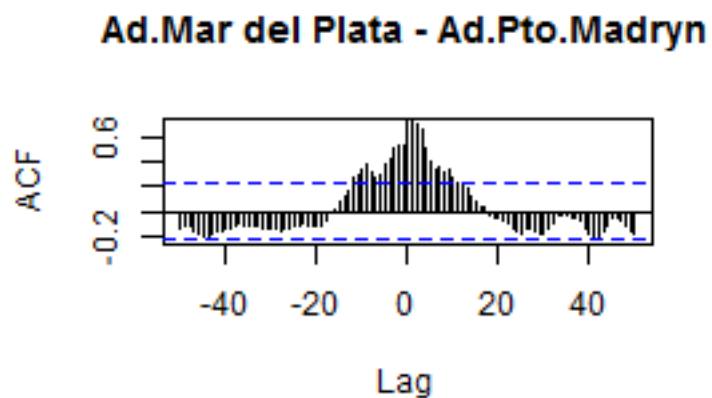


Figure 9: Series mensuales

Ahora consideramos series de seis años y volvemos a realizar las correlaciones. Primero la correlacion entre la aduana Mar del Plata y las otras cinco aduanas maritimas.



Las tres ciudades cercanas entre ellas

Ad.Mar del Plata - Ad.Ushuaia

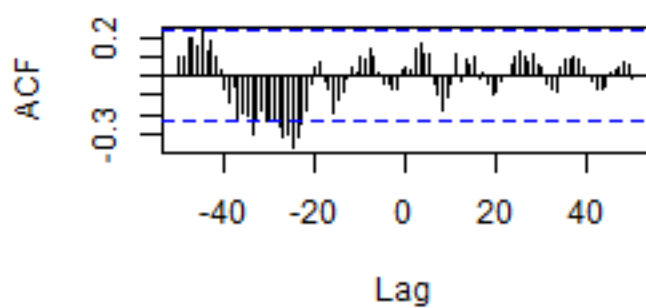
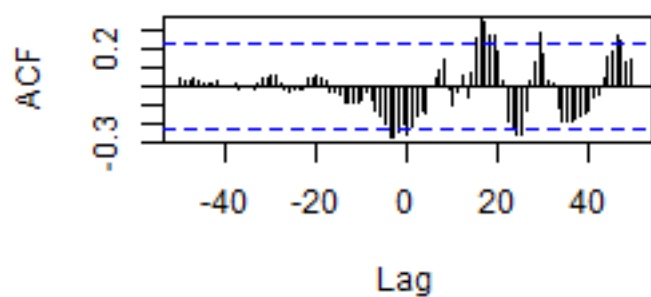
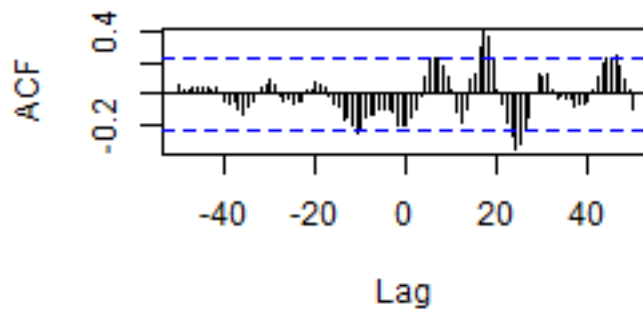


Figure 10: Series mensuales - 6 años

Ad.Cro.Rivadavia - Ad.Caleta Olivia



Ad.Cro.Rivadavia - Ad.Pto.Deseado



Ad.Caleta Olivia - Ad.Pto.Deseado

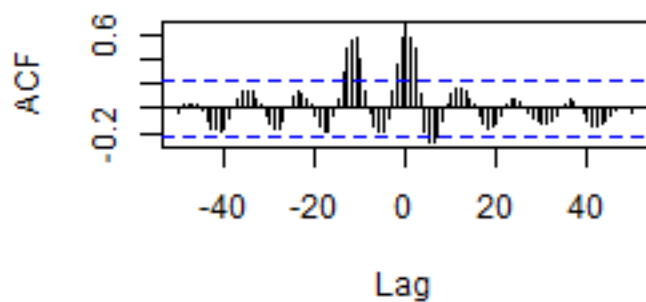


Figure 11: Series mensuales

Y dos ciudades a orillas del rio Parana.

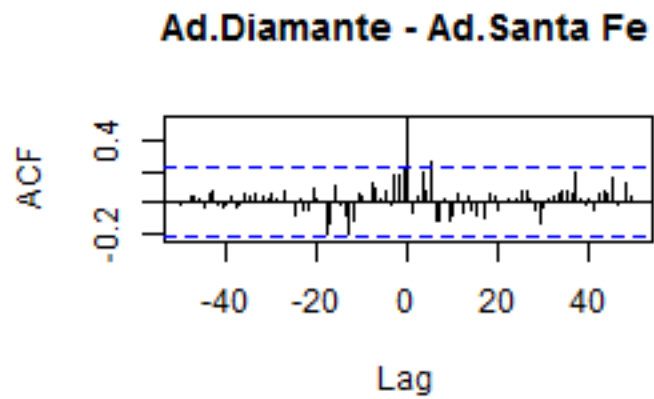


Figure 12: Series mensuales

Generamos scatterplots entre una serie que permanece fija y distintas versiones de otra serie que se mueve desde el valor cero hacia atras mes a mes. Primero se mantiene fija la serie con datos de Mar del Plata (aduana 37) y se mueve mes a mes la serie de Puerto Madryn. Luego se invierten las series.

```
#install.packages("astsa")  
library(astsa)  
lag2.plot (subset.m.47.3$importe, subset.m.37.3$importe, 11)  
lag2.plot (subset.m.37.3$importe, subset.m.47.3$importe, 11)
```

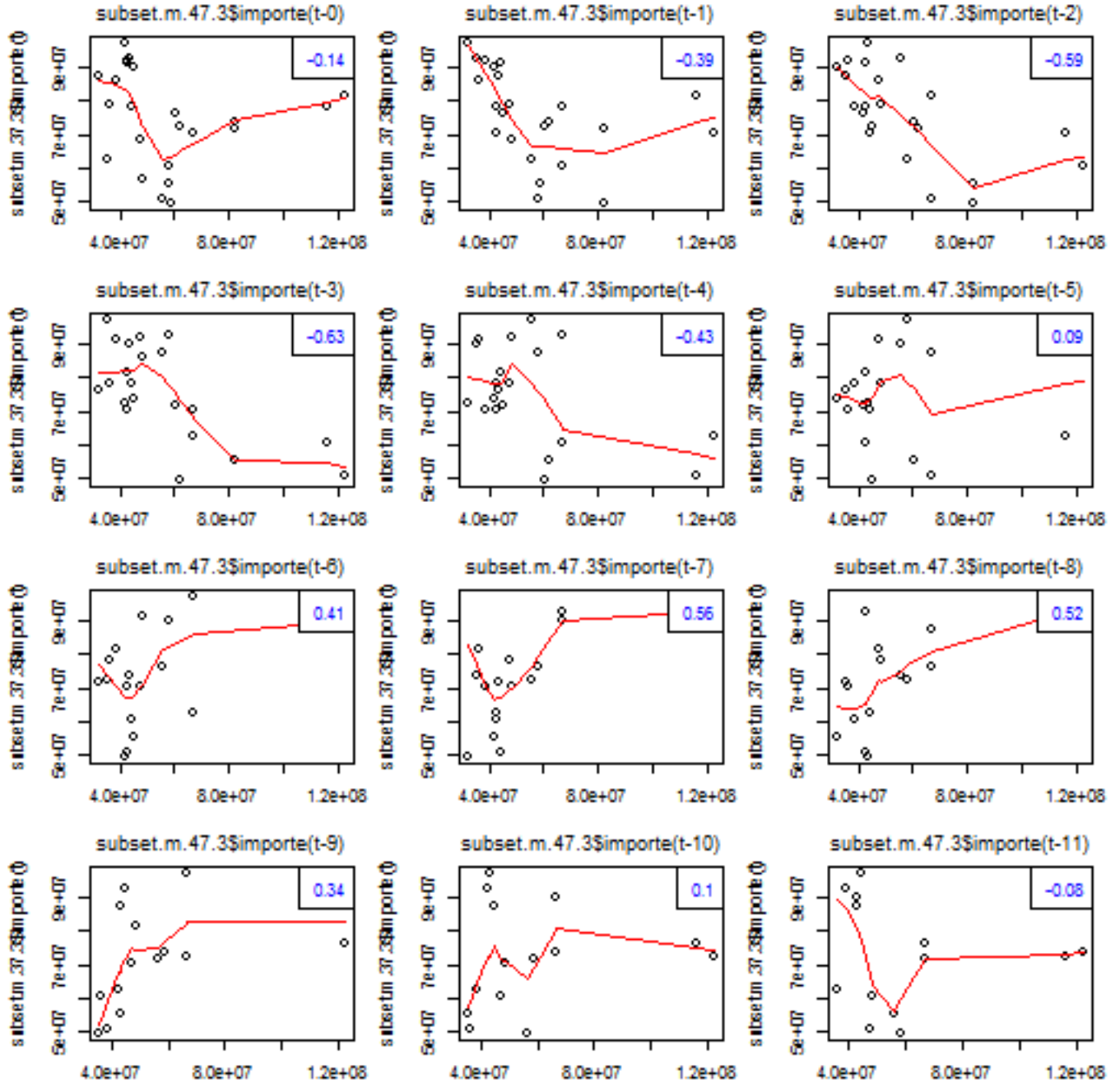


Figure 13: ScatterPlot entre series

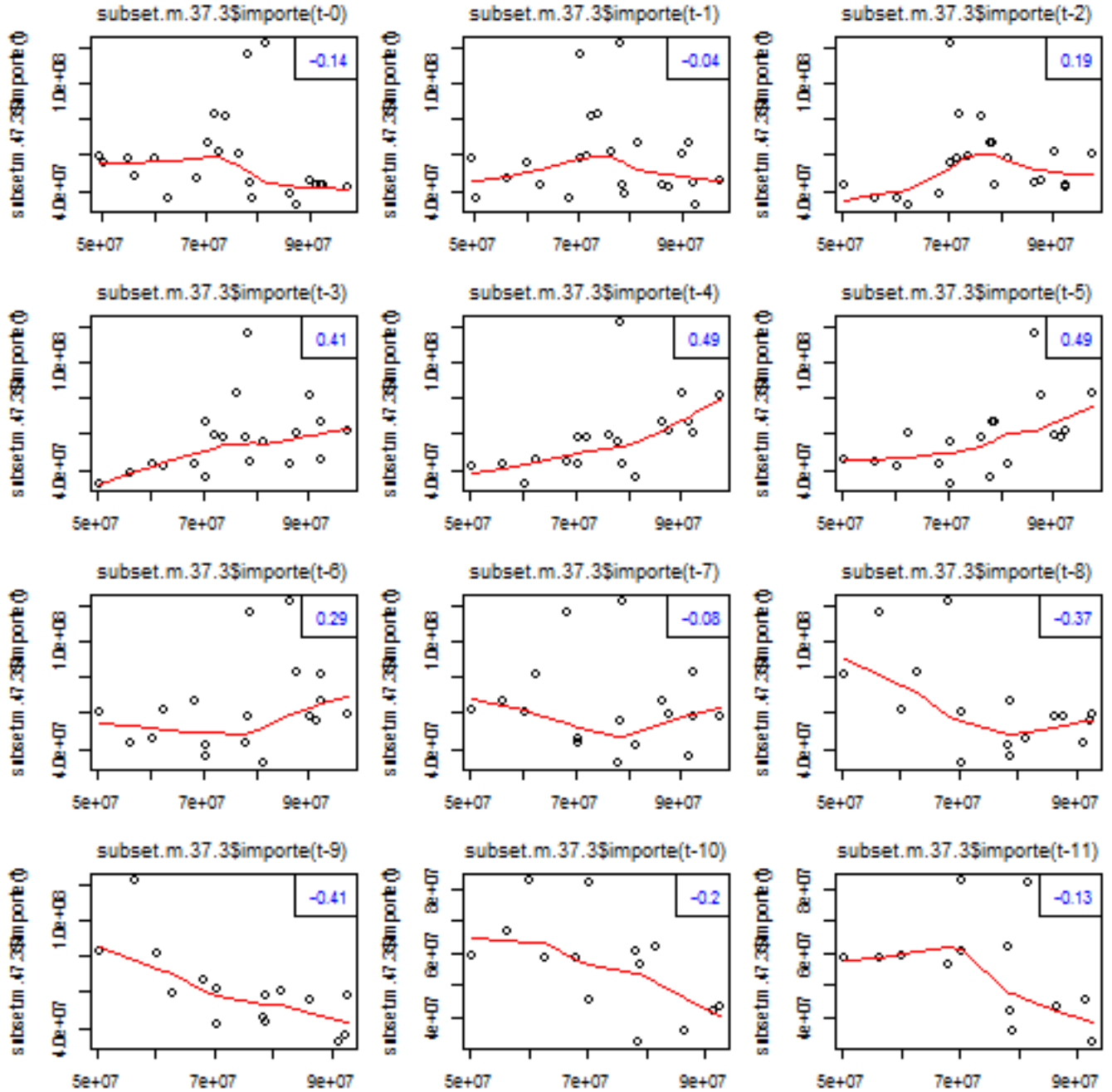


Figure 14: ScatterPlot entre series

Se compara el año 2014 con el año 2015 de las exportaciones de Mar del Plata.

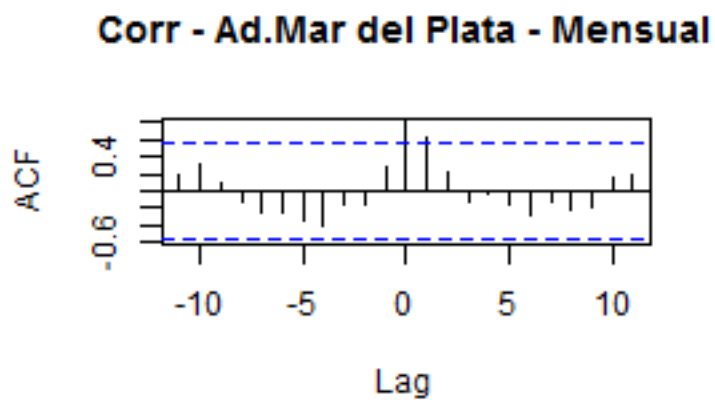


Figure 15: Comparacion entre dos años

Discusión y conclusiones

Bibliografía

Anexos