Examen OPI

## Sección 1

Utilizando la tasa de natalidad de cada año podríamos encontrar una aproximación a la cantidad de bebés que existen al día de hoy con una edad de 0 a 6 meses.

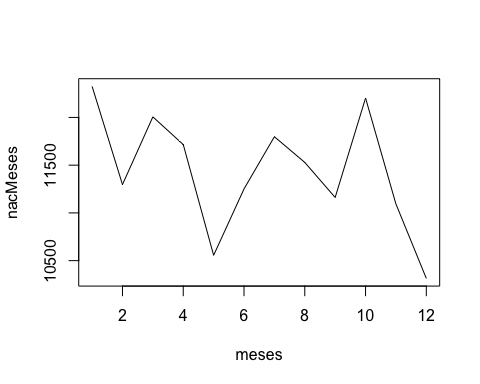
Al estar buscando esa tasa de natalidad, se encontraron las aproximaciones de los índices y datos demográficos para los años del 2010 al 2030.

Estos datos demográficos incluyen la información relacionada a la cantidad de población al inicio y a mediados del año, además de que viene separada por sexo y edad.

Si estamos interesados en la estimación de cantidad de bebés de 0 a 6 meses al día de hoy, podemos considerar la estimación de la población a inicio del 2017 de edad 0. Eso implica que son niños nacidos de enero de 2016 a diciembre de 2016.

pobInicio2017 <- 62809 + 59954

Siendo un total de 122,763 niños y niñas de edad de 0 años.

De la información del 2015, tenemos que en la CDMX hubo un total de 137,249 nacimientos distribuidos de la siguiente forma: 

Supongamos que la proporción de nacimientos en el 2016 es similar al comportamiento ocurrido en el 2015. Es decir, los porcentajes de nacimientos por mes son similares. Estos porcentajes son:

## [1] 0.08980029 0.08231025 0.08747605 0.08536310 0.07691131 0.08197510  
## [7] 0.08596055 0.08400061 0.08132664 0.08890411 0.08082390 0.07514809

Por lo tanto, si queremos conocer la cantidad de bebés de 0 a 6 meses a inicios del 2017 se tendría que observar la cantidad de bebés nacidos de julio a diciembre.

## [1] "Porcentajes de julio a diciembre:"

## [1] 0.08596055 0.08400061 0.08132664 0.08890411 0.08082390 0.07514809

## [1] "Poblacion distribuida por meses"

## [1] 10552.775 10312.167 9983.902 10914.135 9922.185 9225.405

## [1] "Poblacion de 0 a 6 meses"

## [1] 60910.57

Así, la cantidad de bebés en CDMX de 0 a 6 meses es aproximadamente de 60,910. Para obtener el total de bebés en la delegación Álvaro Obregón supongamos que la proporción de nacimientos es similar en el año 2015 y 2016, además de que estaremos suponiendo que los bebés nacidos en una delegación se quedan a vivir en la misma delegación al menos por 6 meses.

## [1] "Porcentajes por delegacion en 2015"

## [1] 0.04386189 0.05487836 0.02656486 0.12678417 0.04241197 0.16579356  
## [7] 0.02457577 0.01879067 0.07835394 0.04403675 0.07151236 0.04005129  
## [13] 0.03997843 0.09504623 0.07621185 0.05114791

## [1] "Porcentaje de la delegacion Alvaro Obregon"

## [1] 0.07835394

## [1] "Cantidad de bebes de 0 a 6 meses en AO"

## [1] 4772.583

Luego, actualmente existe un aproximado de 4772 bebés de 0 a 6 meses de edad en la delegación Álvaro Obregón.

**Fuentes**

<http://www.inegi.org.mx/lib/olap/consulta/general_ver4/MDXQueryDatos.asp?#Regreso&c=23699> <http://www.conapo.gob.mx/es/CONAPO/Proyecciones_Datos>

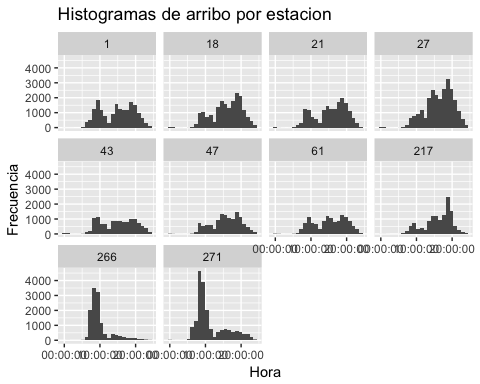
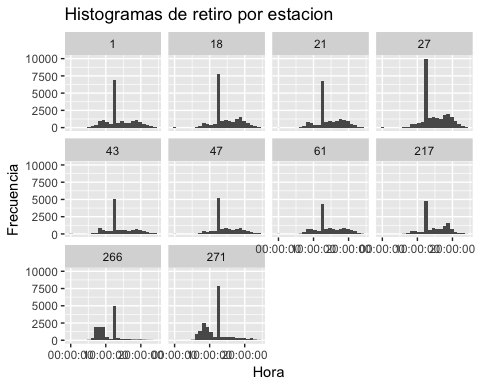
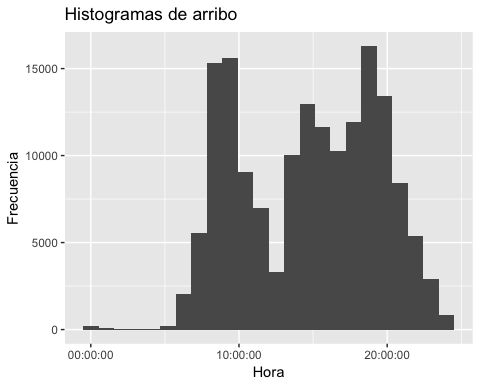
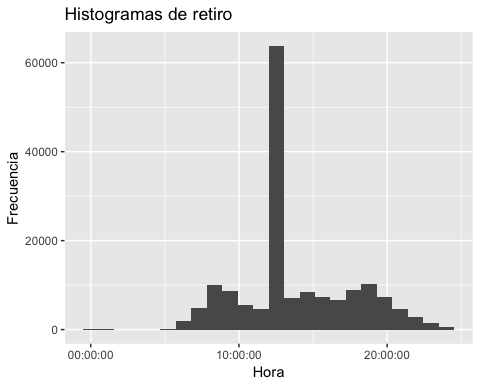
## Sección 2 (ecobici)

**1**

Definamos la afluencia como la cantidad de observacionses que existan en los registros de los tres meses. Esto es, a mayor cantidad de observaciones de una estación, significa que la afluencia en esa estación es mayor. Así, las 10 estaciones con mayor afluencia son

## estacion retiros arribos afluencia  
## 1 27 26067 25785 51852  
## 2 18 19651 19236 38887  
## 3 1 18257 19069 37326  
## 4 271 19747 15807 35554  
## 5 21 17582 15734 33316  
## 6 266 13092 19647 32739  
## 7 61 13570 16045 29615  
## 8 217 13400 15512 28912  
## 9 43 12914 15745 28659  
## 10 47 13106 14319 27425

Si a partir de estas 10 estaciones más frecuentes calculamos un histograma sobre las horas de retiro de una bicicleta, y un histograma sobre las horas de arribo de una bicileta, los horarios más comunes están al rededor de las 10 de la mañana y el medio día, y cercano a las 8 de la noche.



Las horas de mayor frecuencia de uso coincide con las horas de entrada y salida de las oficinas, que son además las zonas donde usualmente están ubicadas las estaciones de ecobici.

**2a**

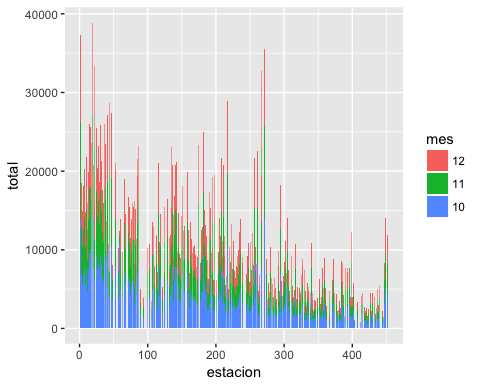
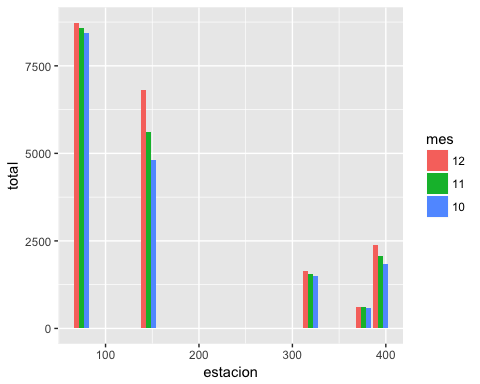
Se puede observar una tendencia al alta en 5 estaciones. Estas estaciones cumplen con el criterio de que el uso en diciembre resultó mayor que en noviembre, y que en noviembre fue mayor que en octubre.

## estacion dic nov oct  
## 1 74 8735 8587 8449  
## 2 146 6817 5620 4810  
## 3 319 1649 1552 1499  
## 4 376 615 600 588  
## 5 394 2387 2066 1840

**2b**

La categorización de las estaciones siempre se puede hacer. Luego es importante verificar que la categorización sea útil, pero siempre se puede establecer alguna. En este caso, para las tendencias de uso, siemplemente se podrían categorizar como "en aumento", "en decrimento" o en "uso constante"; realmente no hay muchas opciones o combinaciones para posibles respuestas o clasificaciones en función de las tendencias de uso de las estaciones de ecobici.

**2c**

La gráfica de las tendencias de uso para las estaciones que van al alta 

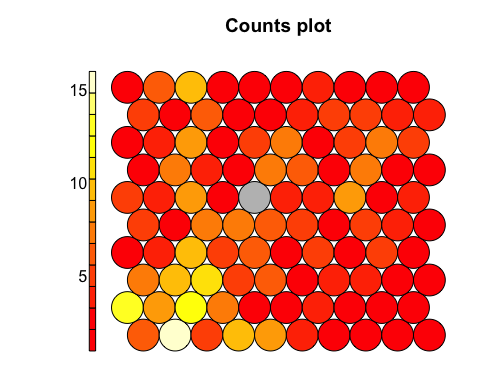
**4**

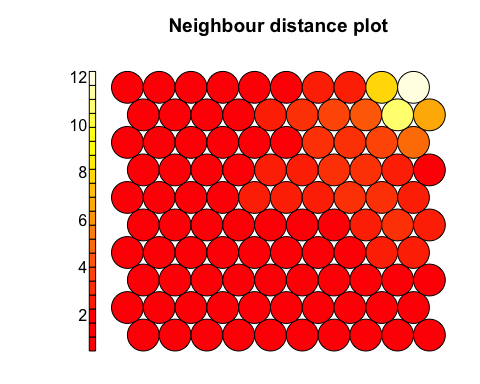
El método de aprendizaje no supervisado que se utilizó para esta parte del examen es "Self Organizing Maps", o también conocida como "Mapeos de Kohonen". Decidí utilizarlos porque es lo que en gral. me ha dado buenos resultados, y técnicamente la ventaja que ofrecen es la conservación de la topología del espacio original, es decir, el resultado del mapeo mantiene la cercanía entre elementos que en el espacio original estaban cercanos.

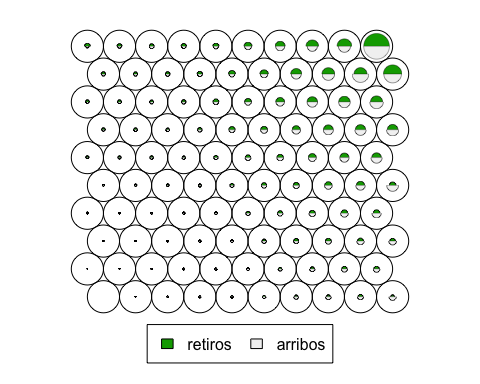
El objetivo "principal" de estos mapeos es representar la información de "alta dimensión" en una dimensionalidad menor, preferentemente en un espacio de dos dimensiones.

La aplicación de SOM a los datos de las estaciones en función de las entradas y las salidas ofrece los siguientes resultados.

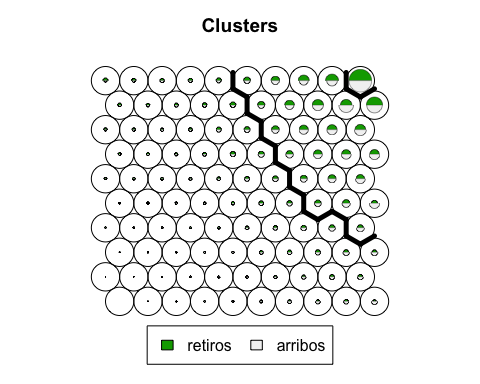
La distribución de elementos en la malla calculada:



El mapeo de disimilitud entre elementos del espacio original es el siguiente: 

La distribución de elementos es: 

Sobre esta distribución podemos hacer 3 grupos para obtener 3 perfiles de uso. Se puede observar en qué consisten estos tres perfiles utilizando esta gráfica:



Los perfiles detectados son básicamente: "de uso intenso", "de uso medio", y "de uso bajo". Las proporciones entre las entradas y las salidas se mantienen similares, por lo tanto la diferencia ocurre en la cantidad de viajes y frecuencia de uso del equipo.