2. Análisis exploratorio.

Considero que las respuestas b) y d) tiene diferentes respuesta, depende del enfoque en el cuál se trate de responder o el objetivo del estudio. La primera manera por la cuál se puede abordar es considerar todos los precios mínimos de los productos de la canasta básica, sin tener en cuenta la distribución geoespacial de la tienda (registros). Pero a mi parecer esto puede tener mucho sesgo, debido a que no se toma encuenta la distribución geospacial, puede que sea la ciudad o estado con el precio de la canasta básica más barata o más cara del país pero no se sabe si los precios mínimos estan muy alejados entres si. El segundo enfoque que se me ocurrió es considerar un promedio del los productos de la canasta básica en esa/ese ciudad/estado, y determinar así un total con la suma de todos los promedios de los productos.

```
#Cargamos los paquetes para los datos
library(sparklyr)
library(tidyverse)
#Configuramos Spark.
config = spark_config()
config$`sparklyr.shell.driver-memory` <- "4G"
config$`sparklyr.shell.executor-memory` <- "4G"
config$`spark.yarn.executor.memoryOverhead` <- "512"
#Realizamos la conección y cargamos los datos en Spark.
sc = spark_connect(master = "local", config = config)

#Leemos los datos. el archivo est_canasta son todos los estados con los 18 productos.
setwd("~/Documentos/Servicio_social/Examen intermedio/")
est_canasta <- spark_read_csv(sc, name = "est", path = "est_canasta.csv/")
```

a) Genera una canasta de productos básicos que te permita comparar los precios geográfica y temporalmente. Justifica tu elección y procedimiento.

Respuesta:

Considerando actualmente que la canasta básica esta constituida de 23 a 40 productos de los cuales en el atributo producto se seleccionaron 18 productos: azucar, frijol, arroz, harina de maiz, aceite, atun, sardina, sal modida de mesa, cafe soluble, harina de trigo, pasata para sopa, avena, lenteja, detergente para ropa, detergente para trastes, jabón de tocador, papel higienico y crema dental. Esta selección resulto de la busqueda de todos los productos en los estados o municipios, es decir, se comprobó que en todo los estados estuvieran los 18 productos especificados. Además, para hacer las comparaciones temporales consideramos los analisis apartir de $2014-06-10$, debido a que antes de esa fecha existían productos sin ningún registros en diversos estados y la fecha en el cuál hubo algún cambio fue el $2016-04-29$.

```
canasta<-c("AZUCAR","FRIJOL","ARROZ","HARINA DE MAIZ","ACEITE","ATUN","SARDINA","SAL MOLIDA DE MESA","C
           "HARINA DE TRIGO","PASTA PARA SOPA","AVENA","LENTEJA","DETERGENTE P/ROPA","DETERGENTE P/TRAS

est_canasta%>%group_by(estado,producto)%>%summarise(min(fechaRegistro))%>%filter(`min(fechaRegistro)`>"
```

```
## # Source:     spark<?> [?? x 3]
## # Groups:     estado
## # Ordered by: estado
##    estado              producto             `min(fechaRegistro)`
##    <chr>               <chr>                <dttm>
##  1 COL. EDUARDO GUERRA ACEITE               2015-10-08 16:23:38
##  2 COL. EDUARDO GUERRA HARINA DE MAIZ       2015-10-08 16:06:47
##  3 COL. EDUARDO GUERRA SAL MOLIDA DE MESA   2016-03-28 06:00:00
##  4 COL. EDUARDO GUERRA HARINA DE TRIGO      2015-10-08 16:07:07
##  5 COL. EDUARDO GUERRA PASTA PARA SOPA      2015-10-08 16:29:26
##  6 COL. EDUARDO GUERRA JABON DE TOCADOR     2015-10-08 16:48:31
##  7 COL. EDUARDO GUERRA ATUN                 2015-10-08 16:36:59
```

```
##  8 COL. EDUARDO GUERRA DETERGENTE P/TRASTES 2015-10-08 16:56:53
##  9 COL. EDUARDO GUERRA LENTEJA              2015-10-08 16:33:29
## 10 COL. EDUARDO GUERRA AVENA                2015-10-08 16:19:51
## # ... with more rows
```

b) ¿Cuál es la ciudad más cara del país? ¿Cuál es la más barata?

Respuesta:

Se me complico esta pregunta, ¿de todo mexico buscamos la ciudad mas cara y más barata? A nivel ciudad, se me complico debido a que le tenia que asignar a cada ciudad la tienda más cercana en la cuál tuvieran los productos, además de que los resgitros tenían el nombre del municipio y no a nivel ciudad, tenía las coordenadas de la tienda en especifico pero no encontre en internet un shapile con todas las coordenadas de las ciudades de México.

c) ¿Hay algún patrón estacional entre años?

d) ¿Cuál es el estado más caro y en qué mes?

Respuesta:

```
canasta<-c("AZUCAR","FRIJOL","ARROZ","HARINA DE MAIZ","ACEITE","ATUN","SARDINA","SAL MOLIDA DE MESA","CA
           "HARINA DE TRIGO","PASTA PARA SOPA","AVENA","LENTEJA","DETERGENTE P/ROPA","DETERGENTE P/TRAST
meses<-seq.Date(from = as.Date("2014-06-10"),to = as.Date("2016-05-10"),by = "month")

#est_canasta<-all_data%>%filter(producto%in%canasta&estado!="COL. EDUARDO GUERRA")%>%na.omit()
#spark_write_csv(est_canasta,"~/Documentos/Servicio_social/Examen intermedio/est_canasta.csv")

precio_cana<-est_canasta%>%group_by(estado,producto,longitud,latitud)%>%filter(estado!="COL. EDUARDO GU

head(precio_cana%>%distinct(estado,canasta)%>%arrange(desc(canasta)))
```

```
## # A tibble: 6 x 2
##   estado              canasta
##   <chr>                 <dbl>
## 1 BAJA CALIFORNIA SUR    191.
## 2 SAN LUIS POTOSÍ        189.
## 3 MORELOS                189.
## 4 NAYARIT                188.
## 5 CHIAPAS                187.
## 6 CAMPECHE               186.
```

Por lo tanto se observa que en el Baja California Sur tiene los precios más bajos para la canasta. Analizaremos como se comporta durante el tiempo.

e) ¿Cuáles son los principales riesgos de hacer análisis de series de tiempo con estos datos?

Respuesta:

Considero que la serie de tiempo se puede ver afectado por la fecha de registros, ya que no todos los productos básicos considerados no tienen el mismo periodo de actualización entre si, por lo que puede influir al realizar el análisis.

Otra detalle importante es que se calculo el precio minimo de la de los productos básicos, pero no se tomo encuenta si los productos tenían la misma presentación.