

Actividad 8

Lorena Pérez

17/5/2021

Esta actividad tiene que quedar disponible en su repositorio de GitHub con el resto de las actividades y tareas del curso. Asegurate que tanto Federico como yo seamos colaboradoras de tu proyecto Tareas_STAT_NT creado hace dos semanas. Recordar seleccionar en las opciones de proyecto, codificación de código UTF-8. Recordar que para que tengas la última versión de tu repositorio en GitHub debes hacer pull a tu repositorio para tener la copia localmente en tu computadora y no generar inconsistencias y antes de terminar subir tus cambios con `commit` y `push`.

1. Ejercicio 1

Cargar la librería `tidyverse` y con los datos `mpg` usar las funciones que vimos de `dplyr` para transformar los datos de la siguiente forma.

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr 0.3.4
## v tibble 3.0.6       v dplyr 1.0.4
## v tidyr 1.1.2        v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

1. Seleccionar el conjunto de autos del 2008 que tienen `hwy` mayor a 31.

```
# data<-as.data.frame(mpg)
data <- mpg
filter(data, year == 2008 & hwy> 31)

## # A tibble: 6 x 11
##   manufacturer model  displ  year  cyl trans      drv    cty   hwy fl    class
##   <chr>          <chr> <dbl> <int> <int> <chr>   <chr> <int> <int> <chr> <chr>
## 1 honda         civic   1.8  2008     4 manual(~ f      26    34 r    subcom~
## 2 honda         civic   1.8  2008     4 auto(15) f      25    36 r    subcom~
## 3 honda         civic   1.8  2008     4 auto(15) f      24    36 c    subcom~
## 4 nissan         altima  2.5  2008     4 manual(~ f      23    32 r    midsize
## 5 toyota        corol~   1.8  2008     4 manual(~ f      28    37 r    compact
## 6 toyota        corol~   1.8  2008     4 auto(14) f      26    35 r    compact

nrow(filter(data, year == 2008 & hwy> 31)) #Son 6 observaciones
```

```
## [1] 6
```

¿Cuántos autos son? Son 6 observaciones

2. Seleccionar el conjunto de autos que tienen cty igual a 9 ó hwy es mayor a 90. ¿Cuántos autos son?

```
filter(data, cty == 9 | hwy > 90)
```

```
## # A tibble: 5 x 11
##   manufacturer model      displ  year  cyl trans  drv      cty   hwy fl      class
##   <chr>          <chr>    <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
## 1 dodge        dakota pi~   4.7  2008     8 auto(~ 4       9    12 e    pick~
## 2 dodge        durango 4~   4.7  2008     8 auto(~ 4       9    12 e    suv
## 3 dodge        ram 1500 ~   4.7  2008     8 auto(~ 4       9    12 e    pick~
## 4 dodge        ram 1500 ~   4.7  2008     8 manua~ 4       9    12 e    pick~
## 5 jeep         grand che~   4.7  2008     8 auto(~ 4       9    12 e    suv
```

```
nrow(filter(data, cty == 9 | hwy > 90)) #Son 5 observaciones
```

```
## [1] 5
```

3. Selecciona la variable displ y hwy

```
select(data, c("displ", "hwy"))
```

```
## # A tibble: 234 x 2
##   displ  hwy
##   <dbl> <int>
## 1  1.8    29
## 2  1.8    29
## 3  2      31
## 4  2      30
## 5  2.8    26
## 6  2.8    26
## 7  3.1    27
## 8  1.8    26
## 9  1.8    25
## 10 2      28
## # ... with 224 more rows
```

Comentario: También puedes seleccionar sin concatener y sin usar character

4. Selecciona desde la variable trans hasta la variable hwy.

```
select(data, trans:hwy)
```

```
## # A tibble: 234 x 4
##   trans      drv      cty   hwy
##   <chr>    <chr> <int> <int>
## 1 auto(l5)  f      18    29
## 2 manual(m5) f      21    29
## 3 manual(m6) f      20    31
## 4 auto(av)  f      21    30
## 5 auto(l5)  f      16    26
## 6 manual(m5) f      18    26
## 7 auto(av)  f      18    27
## 8 manual(m5) 4      18    26
## 9 auto(l5)  4      16    25
## 10 manual(m6) 4      20    28
## # ... with 224 more rows
```

5. Produce un dataframe: marca, modelo, año, cantidad de cilindros y rendimiento en ciudad. Únicamente

para los autos, toyota, camry.

```
data2<-as.data.frame(select(filter(data,manufacturer=="toyota"|manufacturer=="camry"),c("manufacturer",
```

6. Calcula el rendimiento promedio en ciudad para cada marca (manufacturer). ¿Cuál es el mejor rendimiento en ciudad? y el peor ?

```
prom_rend<-data %>%
group_by(manufacturer) %>%
summarise(prom = mean(cty,na.rm=TRUE) )
arrange(prom_rend,desc(prom))
```

```
## # A tibble: 15 x 2
##   manufacturer prom
##   <chr>         <dbl>
## 1 honda         24.4
## 2 volkswagen    20.9
## 3 subaru        19.3
## 4 hyundai       18.6
## 5 toyota        18.5
## 6 nissan         18.1
## 7 audi          17.6
## 8 pontiac       17
## 9 chevrolet     15
## 10 ford         14
## 11 jeep         13.5
## 12 mercury      13.2
## 13 dodge        13.1
## 14 land rover   11.5
## 15 lincoln      11.3
```

La marca que tiene mayor rendimiento promedio es Honda, mientras que Lincoln es que tiene rendimiento promedio más bajo.

7. Para cada marca: calcula el rendimiento promedio en ciudad, el error estandar de la media, y el rango

```
data %>%
group_by(manufacturer) %>%
summarise(prom = mean(cty,na.rm=TRUE),ee<-sd(cty),rango<-max(cty)-min(cty) )
```

```
## # A tibble: 15 x 4
##   manufacturer prom 'ee <- sd(cty)' 'rango <- max(cty) - min(cty)'
##   * <chr>         <dbl>         <dbl>         <int>
## 1 audi          17.6         1.97           6
## 2 chevrolet     15         2.92          11
## 3 dodge         13.1         2.49           9
## 4 ford          14         1.91           7
## 5 honda         24.4         1.94           7
## 6 hyundai       18.6         1.50           5
## 7 jeep          13.5         2.51           8
## 8 land rover    11.5         0.577          1
## 9 lincoln       11.3         0.577          1
## 10 mercury      13.2         0.5            1
## 11 nissan        18.1         3.43          11
## 12 pontiac      17         1             2
## 13 subaru       19.3         0.914          3
## 14 toyota       18.5         4.05          17
```

```
## 15 volkswagen      20.9      4.56      19
```

8. Mejora en el rendimiento: calcular el rendimiento promedio para cada marca, distinguiendo antes y después de 2004

```
dataa<-filter(data, year< 2004)
rend1<-dataa %>%
group_by(manufacturer) %>%
summarise(pre2004= mean(cty,na.rm=TRUE))
datab<-filter(data, year>= 2004)
rend2<-datab %>%
group_by(manufacturer) %>%
summarise(post2004 = mean(cty,na.rm=TRUE))
rend<-merge(rend1,rend2)
rend
```

```
##      manufacturer pre2004 post2004
## 1          audi 17.11111 18.11111
## 2      chevrolet 15.14286 14.91667
## 3          dodge 13.37500 12.95238
## 4          ford 13.93333 14.10000
## 5          honda 24.80000 24.00000
## 6        hyundai 18.33333 18.87500
## 7          jeep 14.50000 13.16667
## 8    land rover 11.00000 12.00000
## 9        lincoln 11.00000 12.00000
## 10       mercury 13.50000 13.00000
## 11        nissan 17.66667 18.42857
## 12       pontiac 17.00000 17.00000
## 13        subaru 19.00000 19.50000
## 14        toyota 18.15000 19.07143
## 15    volkswagen 21.25000 20.45455
```

Comentario: Correcto resultado pero no la forma de realizarlo revisar la solución.

9. Calcular el rendimiento promedio en carretera (hwy), para 3 marcas seleccionadas aleatoriamente y ordena el resultado de menor a mayor

```
data3<-filter(data,manufacturer==sample(unique(data$manufacturer),3))
rendhwy<-data3 %>%
group_by(manufacturer) %>%
summarise(rend= mean(hwy,na.rm=TRUE))
arrange(rendhwy)
```

```
## # A tibble: 3 x 2
##   manufacturer  rend
##   <chr>        <dbl>
## 1 hyundai      27.2
## 2 jeep        15
## 3 pontiac     27
```

10. Crear una nueva variable que transforme hwy (millas por galón) en litros/100 km. Primero tenés que saber la conversión de galones a litros y de millas a kilómetros.

1 gallón = 3.78541 litros

1 milla = 1.60934 km

```
data4<-mutate(data, newhwy = 235.21/hwy)
```

Comentario: Los tibbles ya son data.frames no es necesario convertirlos. A mejorar: Hay calculos intermedios que no son necesarios, el código debería estar indentado para que sea más legible y no es necesario concatenar al usar select o filter ni usar character (pero si se quiere se puede hacer no hay nada malo en ello)