

Actividad 6

Martín Gabriel Blanco Desvars

1/4/2018

Ejercicio 1

1. Crear una nuevo repositorio en GitHub llamado Actividad 6.
2. Crear un proyecto de RStudio y conectarlo con su nuevo repositorio en GitHub
3. Abrir un archivo de Rmarkdown que compile a pdf con la soluci?n de la Actividad 6, puede usar como base el .Rmd que est? disponible en EVA.
4. Realizar el primer `commit` y `push` con la Actividad 6 en su nuevo repositorio de GitHub (s?lo el .Rmd).

Ejercicio 2

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

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

Para que no salgan los mensajes y warnings en el chunk de código tenés que agregar `'warning=FALSE, message=FALSE'`

```
library(tidyverse)
```

```
new_data <- filter(mpg, year ==2008, hwy > 31)
new_data
```

```
## # 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  1.80  2008     4 manu~ f     26    34 r   subc~
## 2 honda        civic  1.80  2008     4 auto~ f     25    36 r   subc~
## 3 honda        civic  1.80  2008     4 auto~ f     24    36 c   subc~
## 4 nissan        alti~  2.50  2008     4 manu~ f     23    32 r   mids~
## 5 toyota        coro~  1.80  2008     4 manu~ f     28    37 r   comp~
## 6 toyota        coro~  1.80  2008     4 auto~ f     26    35 r   comp~
```

Quedan 6 autos

Bien!

2. Seleccionar el conjutno de autos que tienen `cty` igual a 9 ? `hwy` es mayor a 90. ?Cuántos autos son?

```
new_data <- filter(mpg, cty == 9 | hwy > 90)
new_data
```

```
## # 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        dako~  4.70  2008     8 auto~ 4     9    12 e   pick~
## 2 dodge        dura~  4.70  2008     8 auto~ 4     9    12 e   suv
```

```
## 3 dodge      ram ~ 4.70 2008      8 auto~ 4          9      12 e      pick~
## 4 dodge      ram ~ 4.70 2008      8 manu~ 4          9      12 e      pick~
## 5 jeep       gran~ 4.70 2008      8 auto~ 4          9      12 e      suv
```

Bien!

Quedan 5 autos

3. Selecciona la variable `displ` y `hwy`

```
new_data <- select(mpg, displ, hwy)
new_data
```

```
## # A tibble: 234 x 2
##   displ  hwy
##   <dbl> <int>
## 1  1.80    29
## 2  1.80    29
## 3  2.00    31
## 4  2.00    30
## 5  2.80    26
## 6  2.80    26
## 7  3.10    27
## 8  1.80    26
## 9  1.80    25
## 10 2.00    28
## # ... with 224 more rows
```

Bien!

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

```
new_data <- select(mpg, trans:hwy)
new_data
```

```
## # 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
```

Bien!

5. 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 galón = 3.78541 litros

1 milla = 1.60934 km

```
new_data <- mutate(mpg, hw= hwy*0.42514)
new_data
```

```
## # A tibble: 234 x 12
##   manufacturer model   displ  year  cyl trans  drv    cty   hwy fl
##   <chr>          <chr>   <dbl> <int> <int> <chr>  <chr> <int> <int> <chr>
## 1 audi          a4       1.80  1999    4 auto(l~ f    18    29 p
## 2 audi          a4       1.80  1999    4 manual~ f    21    29 p
## 3 audi          a4       2.00  2008    4 manual~ f    20    31 p
## 4 audi          a4       2.00  2008    4 auto(a~ f    21    30 p
## 5 audi          a4       2.80  1999    6 auto(l~ f    16    26 p
## 6 audi          a4       2.80  1999    6 manual~ f    18    26 p
## 7 audi          a4       3.10  2008    6 auto(a~ f    18    27 p
## 8 audi          a4 quat~ 1.80  1999    4 manual~ 4    18    26 p
## 9 audi          a4 quat~ 1.80  1999    4 auto(l~ 4    16    25 p
## 10 audi         a4 quat~ 2.00  2008    4 manual~ 4    20    28 p
## # ... with 224 more rows, and 2 more variables: class <chr>, hw <dbl>
```

```
x<- select(new_data,hwy, hw)
```

```
x
```

```
## # A tibble: 234 x 2
##   hwy  hw
##   <int> <dbl>
## 1    29 12.3
## 2    29 12.3
## 3    31 13.2
## 4    30 12.8
## 5    26 11.1
## 6    26 11.1
## 7    27 11.5
## 8    26 11.1
## 9    25 10.6
## 10   28 11.9
## # ... with 224 more rows
```

Bien

#Ejercicio 3

Para esta actividad vamos a utilizar los datos happy que se encuentran en la librería productplots. Con estos datos vamos a investigar algunas preguntas sobre que cosas afectan la felicidad de la gente. Para responder estas preguntas utilizaremos dplyr

1. Encontrar el porcentaje para los distintos niveles de felicidad para hombres y mujeres a lo largo de todos los años

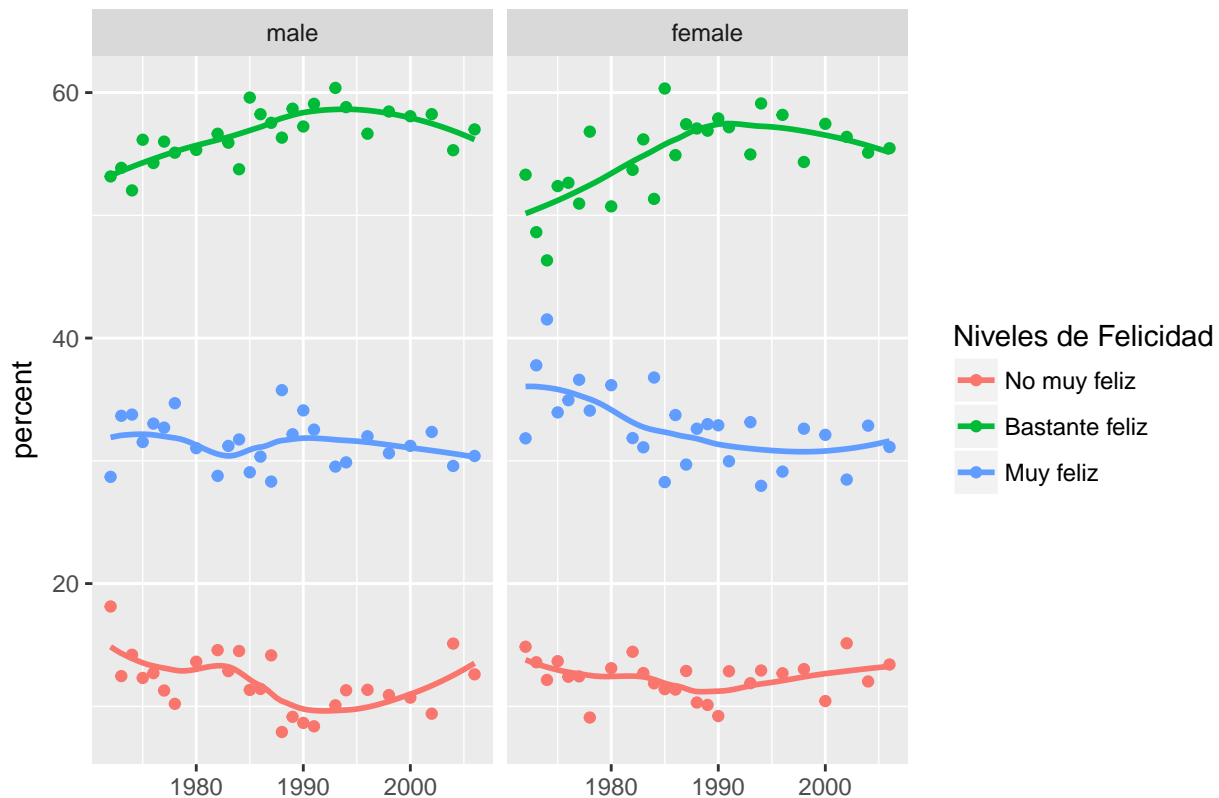
```
library(productplots)
new_data <- filter(happy, !is.na(happy)) %>% group_by(sex,happy,year) %>% summarise(cont=n()) %>% arrange
```

Bien

2. Graficar los porcentajes en un gráfico de dispersión con year en el eje x y porcentaje en el eje y. Tendrá que haber un total de seis puntos para cada año, colorear los puntos según happy y remover los NA en happy usando filter.

```
new_data %>% ggplot(aes(x= year , y= percent , colour= happy)) + geom_point() + facet_wrap(~sex) + geom_smooth()
```

Gráfico de Dispersión: Niveles de felicidad a lo largo del tiempo ()



Bien, aquí no pedimos separar por sexo y por eso deberán haber 6 puntos para cada año.

3. ¿Cómo se ve afectada la felicidad según el género?

El comportamiento de la variable felicidad es semejante en ambos sexos, sin embargo al nivel “muy feliz” el porcentaje de mujeres es mayor.

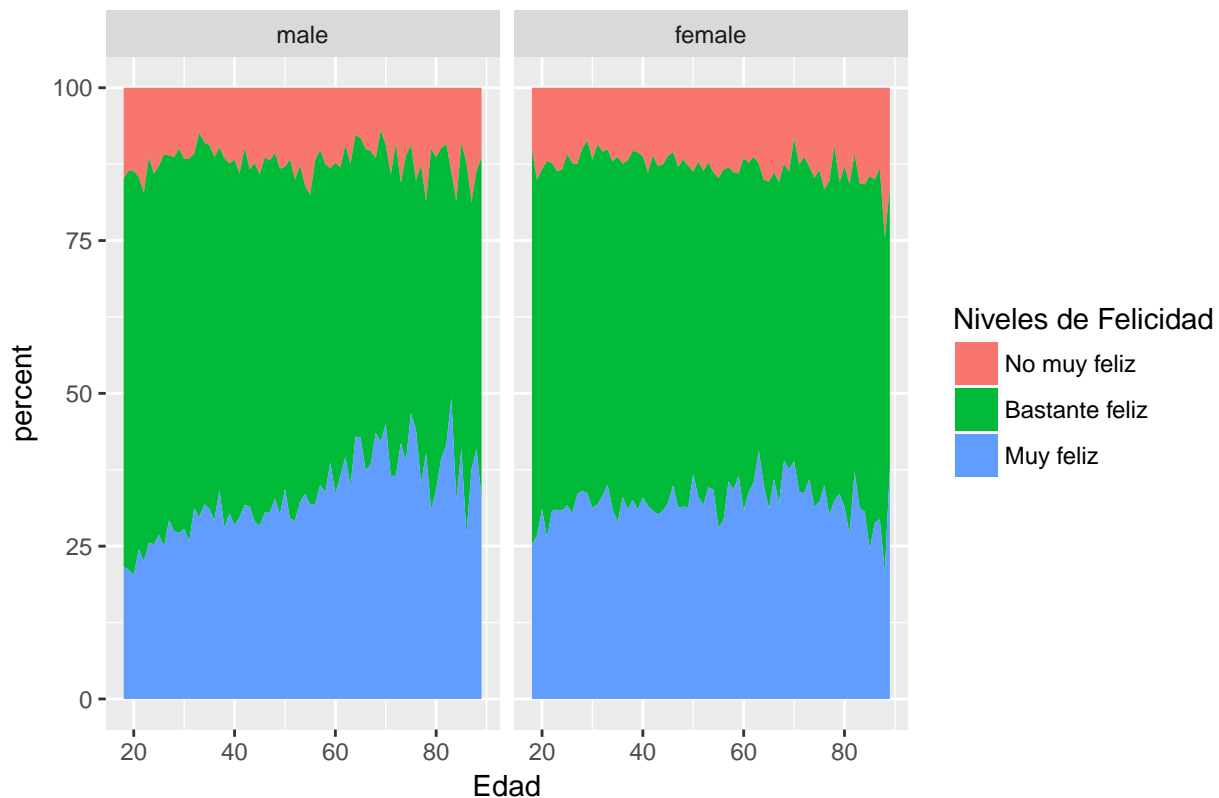
Mirá la solución, estás usando el plot previo para esta respuesta pero se puede ver sin diferenciar por año

4. ¿Cómo se ve afectada la felicidad según la edad? ¿Existe una diferencia entre hombres y mujeres?

```
filter(happy, !is.na(happy)) %>% group_by(sex, happy, age) %>% summarise(count=n()) %>% group_by(age, sex) %>%
```

```
## Warning: Removed 6 rows containing missing values (position_stack).
```

Gráfico de área: Niveles de felicidad según sexo y edad



El comportamiento de la variable felicidad es muy semejante a excepción de una tendencia positiva a lo largo del tiempo para el nivel muy feliz en la población masculina.

Bien

5. ¿Y si analizamos por estado civil y situación financiera qué encontramos?

```
library(ggmosaic)
```

```
##
```

```
## Attaching package: 'ggmosaic'
```

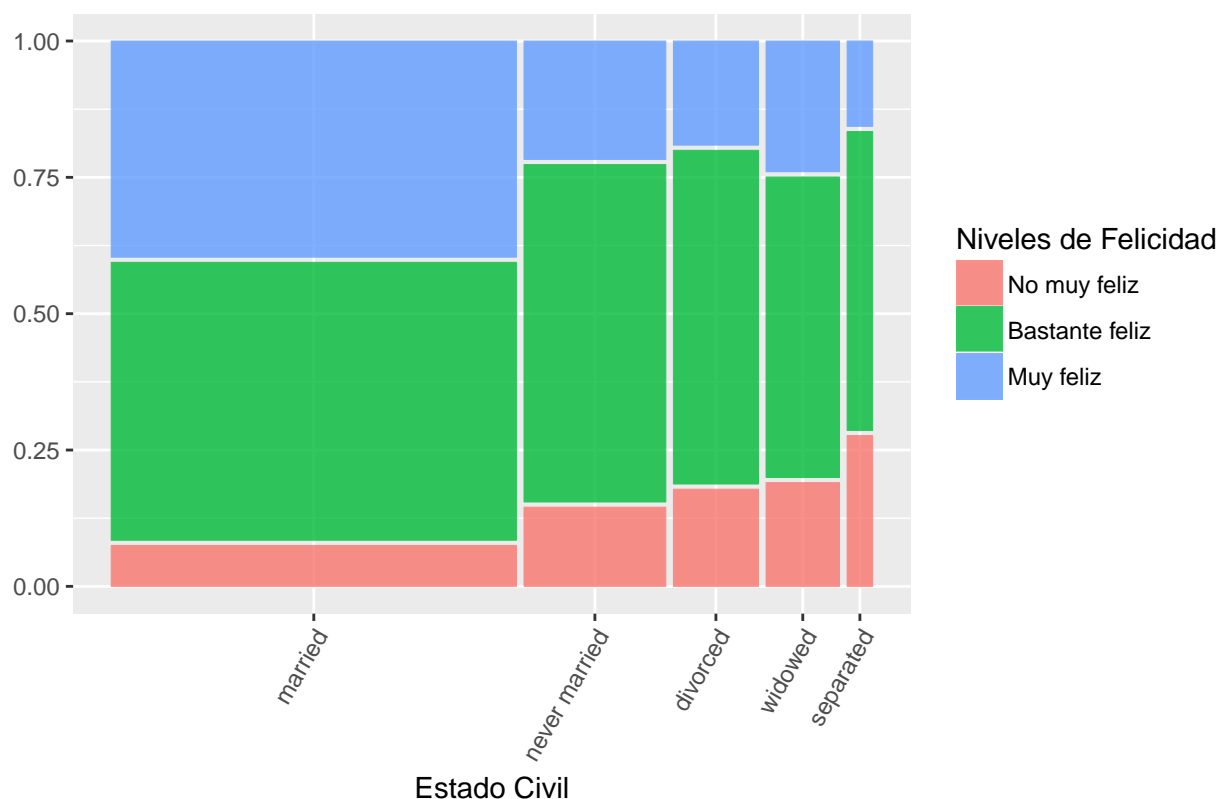
```
## The following objects are masked from 'package:productplots':
```

```
##
```

```
## ddecker, hspine, mosaic, prodcalc, spine, vspine
```

```
happy %>% filter(!is.na(happy)) %>% group_by(happy, marital) %>% arrange(marital) %>% group_by(marital, happy) %>%  
  ggplot() +  
  geom_mosaic(aes( weight=cont, x=product(happy, marital), colour=happy, fill=happy), na.rm=TRUE,)+ labs(x="Marital Status", y="Happiness Level")
```

Gráfico de Mosaicos: Niveles de felicidad según estado civil

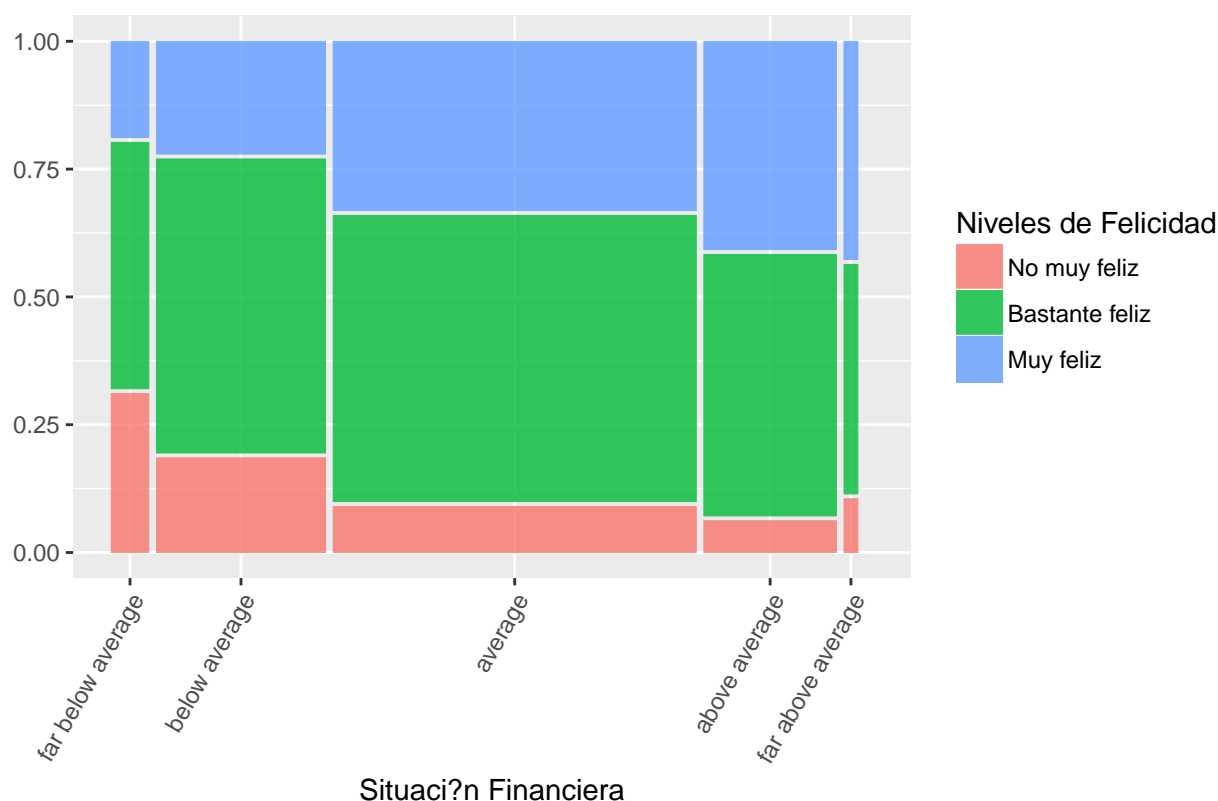


La población de personas casadas cuenta con la mayor proporción de individuos muy felices contrastando con la población separada que cuenta con proporción de individuos no muy felices.

Bien en incluir mosaicos, pero podrías comentar algo sobre la diferencia en tamaño de muestra para los distintos niveles de estado civil y que usaste mosaicos

```
happy %>% filter(!is.na(happy)) %>% group_by(happy,finrela) %>% arrange(finrela)%>% group_by(finrela,happy)
ggplot() +
  geom_mosaic(aes( weight=cont,x =product(happy,finrela), colour=happy, fill=happy))+ scale_fill_discrete(2)
```

Gráfico de Mosaicos: Niveles de felicidad segùn situaciòn financiera



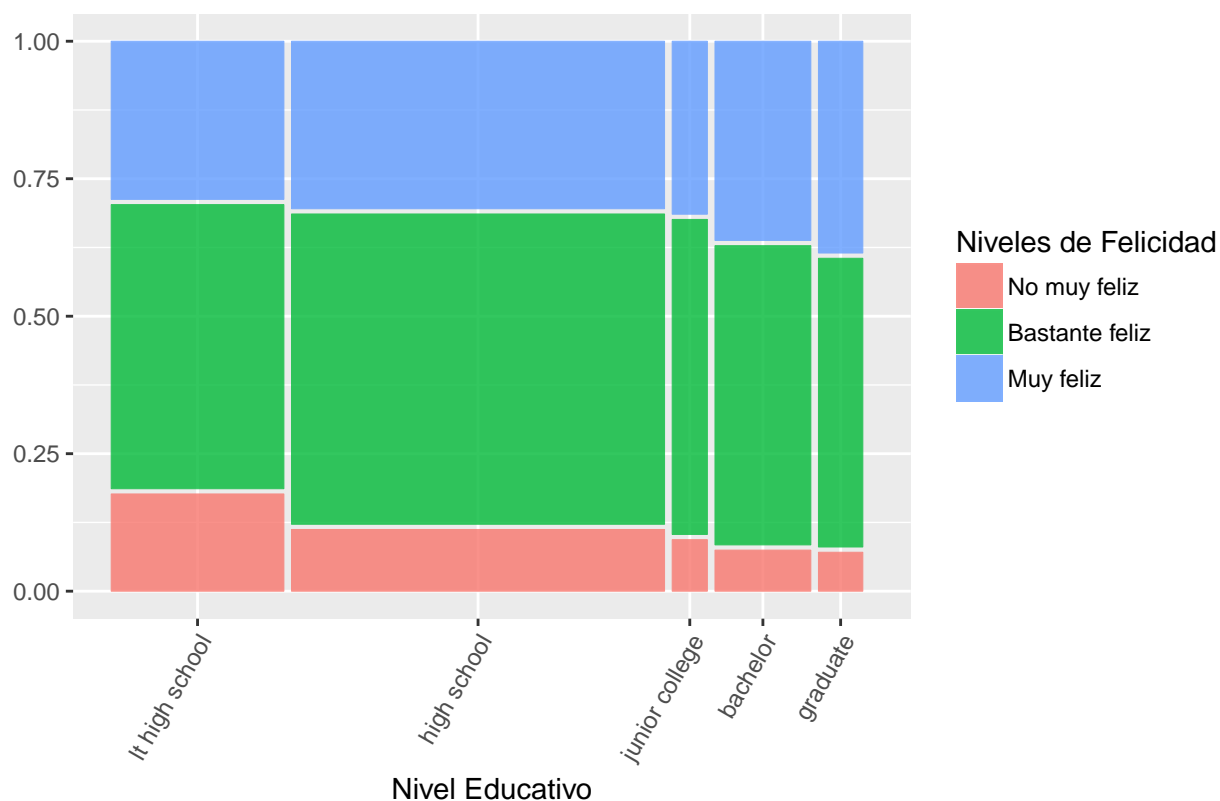
Se observa una tendencia positiva a medida que mejora la situaciòn financiera tambièn aumenta la proporciòn de individuos muy felices, disminuyendo la proporciòn de individuos no muy felices.

Bien !

6. En relaciòn a la variable educaciòn ¿Què se puede ver?

```
happy %>% filter(!is.na(happy)) %>% group_by(happy, degree) %>% arrange(degree) %>% group_by(degree, happy)
ggplot() +
  geom_mosaic(aes( weight=cont, x =product(happy, degree), colour=happy, fill=happy)) + scale_fill_discrete()
```

Gráfico de Mosaicos: Niveles de felicidad segùn nivel educativo



Similar a la situaciòn financiera medida que mejora el nivel educativo tambièn aumenta la proporciòn de individuos muy felices, disminuyendo la proporciòn de individuos no muy felices.

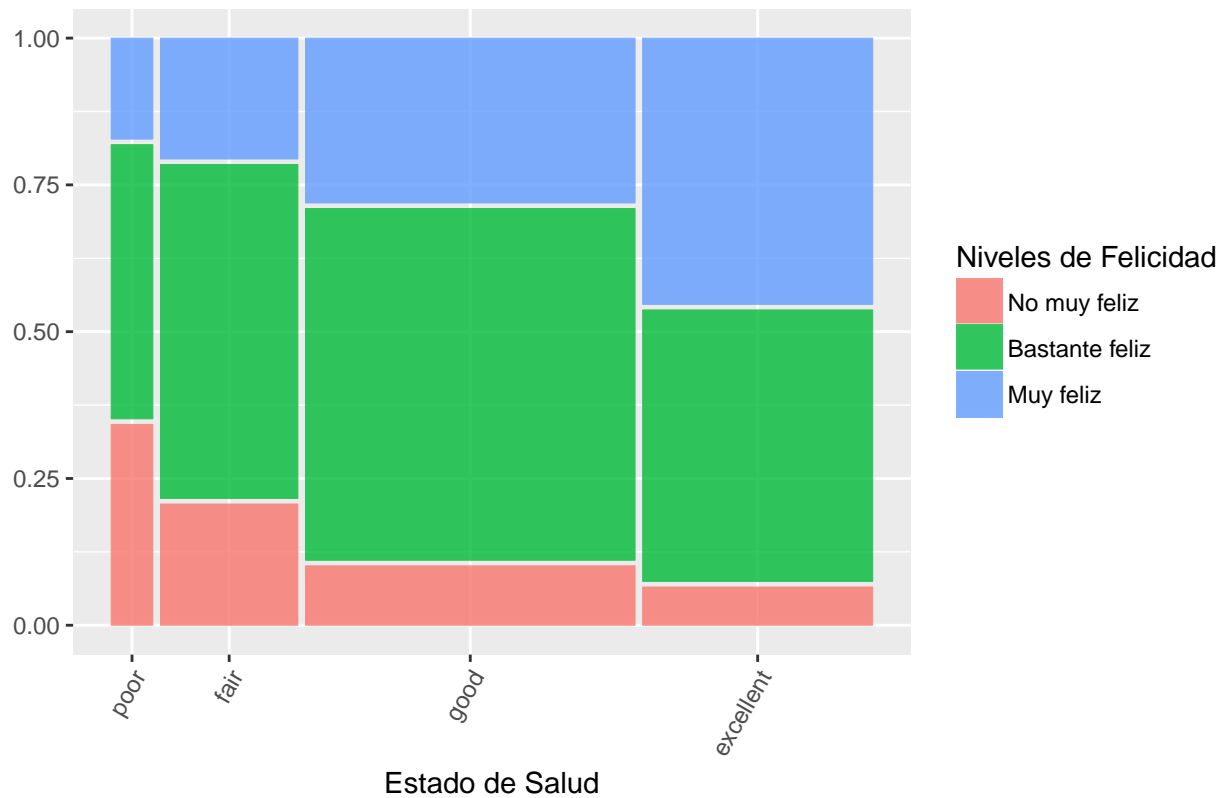
7. ¿Y respecto a salud?

```
levels(happy$health)
```

```
## [1] "poor"      "fair"      "good"      "excellent"
```

```
happy %>% filter(!is.na(happy)) %>% group_by(happy,health) %>% arrange(health) %>% group_by(health,happy) %>%
  ggplot() +
  geom_mosaic(aes( weight=cont, x=product(happy,health), colour=happy, fill=happy))+ scale_fill_discrete(1)
```


Gráfico de Mosaicos: Niveles de felicidad según estado de salud



De los individuos que poseen excelente estado de salud aproximadamente el 47% es muy feliz mientras que el 38% de los individuos en condición de salud pobre se encuentran no muy felices.

Buen trabajo, Puntos 9/10