

# Guia 17

Agustín Muñoz González

17/6/2020

## Preparamos el entorno

```
rm(list=ls())
library(ggplot2)
library(tidyr)
library(gganimate)
```

A lo largo de esta clase vamos a trabajar con diferentes estimadores y sus correspondientes estimaciones, procurando afianzar la idea de que un estimador es una variable aleatoria con su propia distribución, mientras que una estimación es apenas una realización del estimador utilizando los datos observados. Para acceder a los datos que le permitirán resolver los siguientes ejercicios haga click [aquí](#). Recuerde utilizar su número de libreta o cualquier otro número que lo identifique, para que pueda trabajar con SUS datos, siempre que lo desee. Complete el este documento con los resultados que se solicitan.

## 1. Emulando un laboratorio

### Mediciones de gas - Equipo 1

Considere  $n = 100$  datos obtenidos al utilizar el equipo 1.

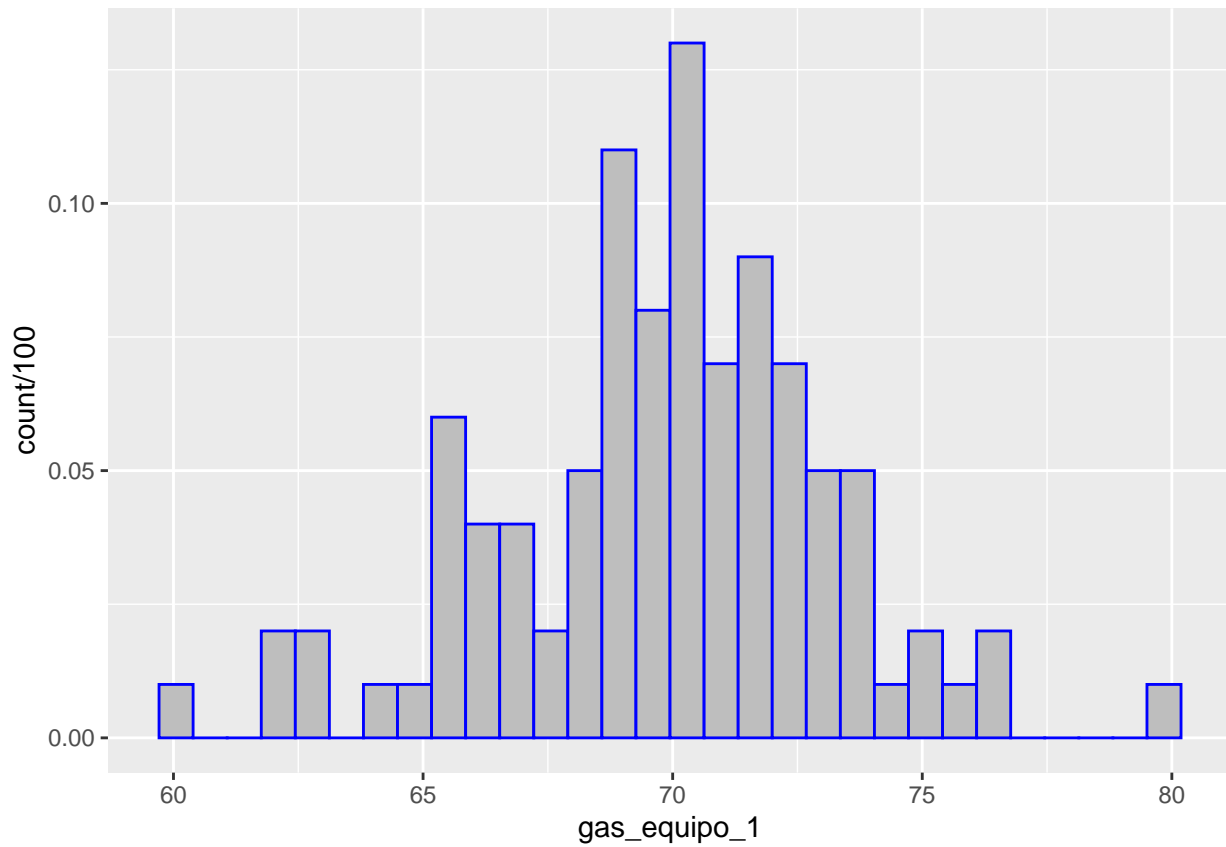
1. Realice un histograma.
2. Calcule el promedio de los datos.
3. Calcule el percentil 0.9 de los datos.
4. Estime la probabilidad de que una medición realizada con este equipo diste de 70 en más de 2 unidades.
5. Repita los ítems anteriores utilizando ahora los primeros  $n = 5$  y  $n = 30$  datos.

Resolución:

1. Realice un histograma.

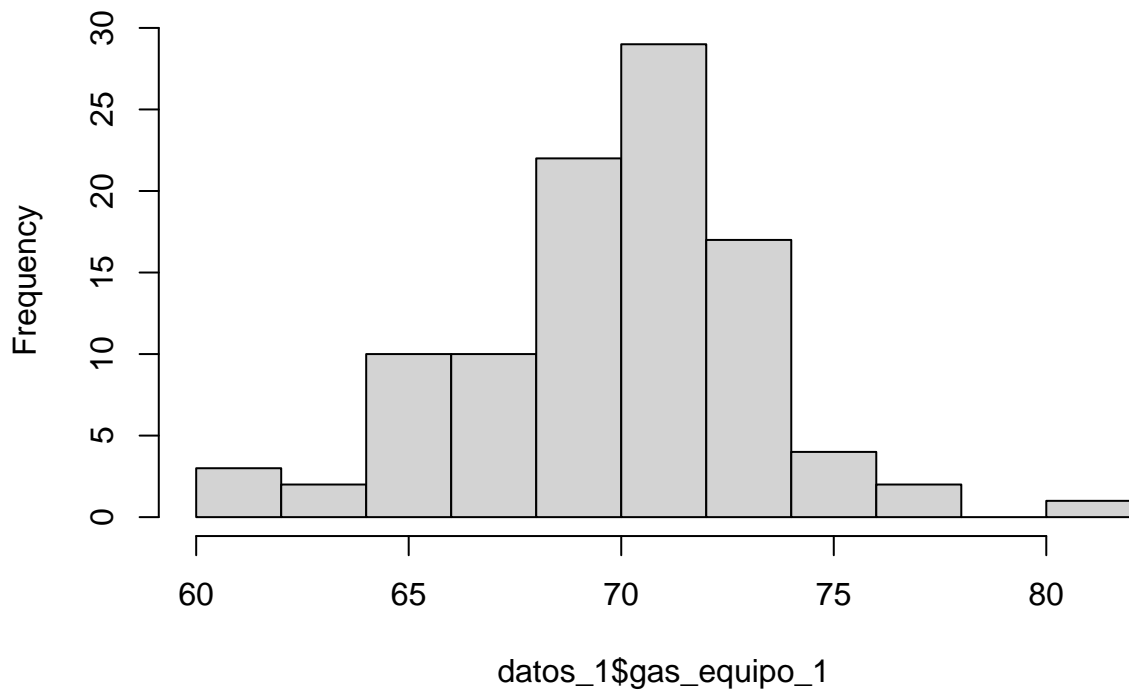
```
datos_1=read.csv('datos_gas1_n_100.csv', header=TRUE)
datos_1 %>%
  ggplot()+
  geom_histogram(aes(x=gas_equipo_1,y=stat(count)/100),
                 color='blue',fill='grey')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
hist(datos_1$gas_equipo_1)
```

**Histogram of datos\_1\$gas\_equipo\_1**



2. Calcule el promedio de los datos.

```
mean(datos_1$gas_equipo_1)
```

```
## [1] 69.8413
```

3. Calcule el percentil 0.9 de los datos.

```
quantile(datos_1$gas_equipo_1,0.9)
```

```
## 90%
```

```
## 73.633
```

4. Estime la probabilidad de que una medición realizada con este equipo diste de 70 en más de 2 unidades.

Queremos

$$P(|datos_1 - 70| > 2).$$

Y este valor sabemos que (como toda proba) lo aproximamos con la frecuencia relativa de datos que cumplen eso. Es decir, de todos los datos me quedo con los que distan de 70 a mas de 2 unidades y me fijo la frec relativa del 1, i.e. la media (porque son 0s y 1s -> mean=freq rel).

```
mean(abs(datos_1$gas_equipo_1-70)>2)
```

```
## [1] 0.49
```

Es decir, hay un 50% de probabilidades de que un valor diste de 70 en mas de 2 unidades. Con lo cual podemos pensar que 70 es el “punto medio” de los datos.

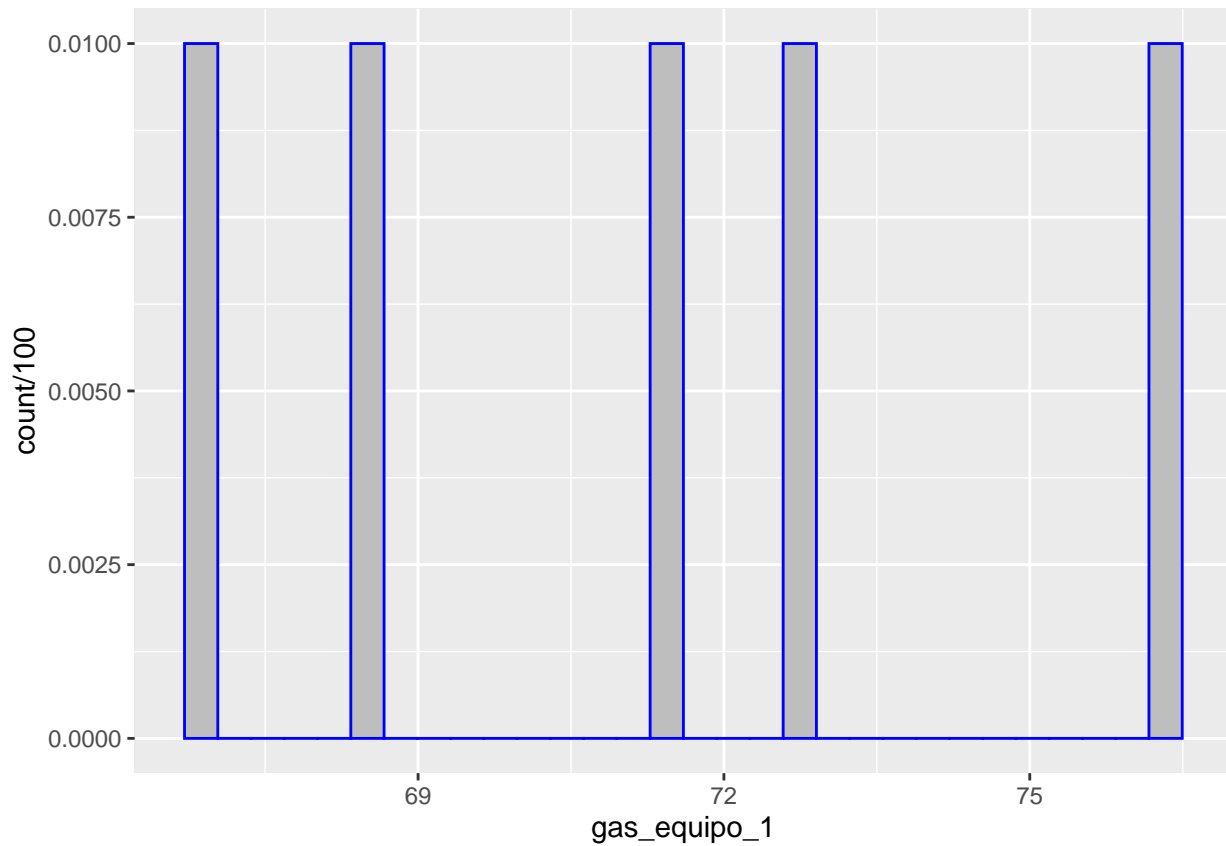
Repetimos para n=5 y n=30.

n=5

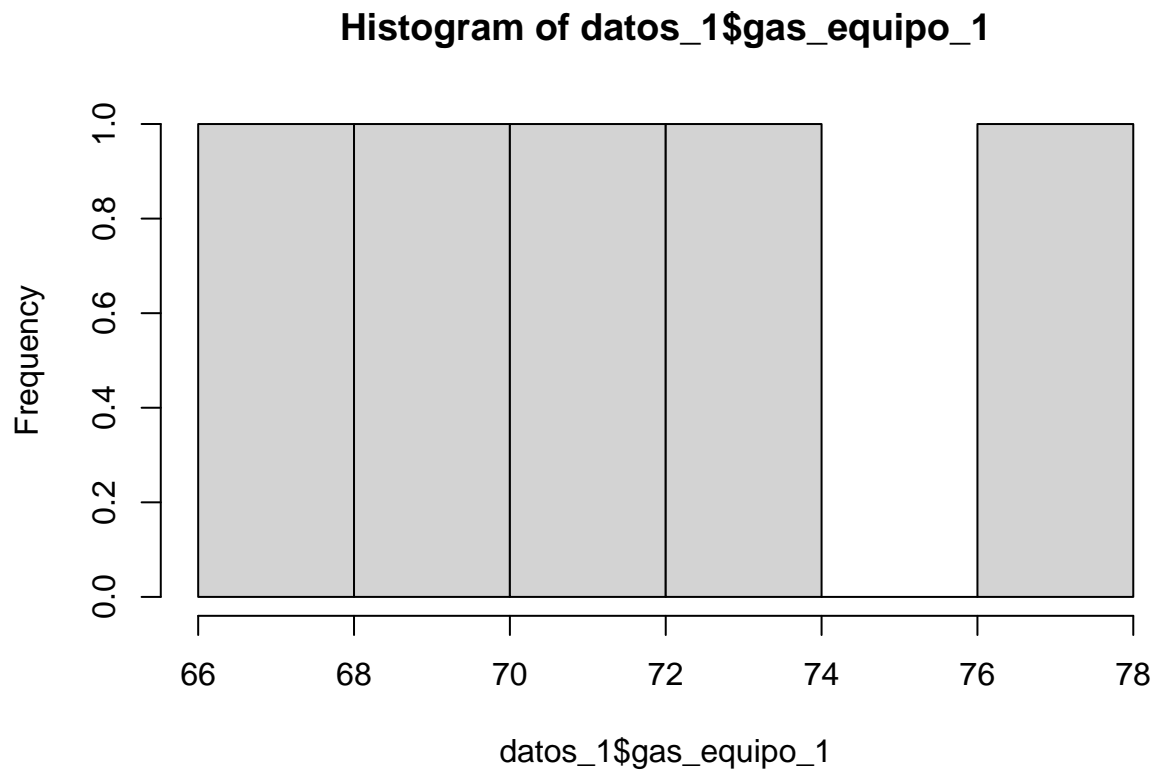
1. Realice un histograma.

```
n=5
datos_1=read.csv('datos_gas1_n_100.csv', header=TRUE)
datos_1=data.frame('gas_equipo_1'=datos_1[1:n,])
datos_1 %>%
  ggplot()+
  geom_histogram(aes(x=gas_equipo_1,y=stat(count)/100),
                 color='blue',fill='grey')

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
hist(datos_1$gas_equipo_1)
```



2. Calcule el promedio de los datos.

```
mean(datos_1$gas_equipo_1)
```

```
## [1] 71.178
```

3. Calcule el percentil 0.9 de los datos.

```
quantile(datos_1$gas_equipo_1,0.9)
```

```
## 90%
```

```
## 74.954
```

4. Estime la probabilidad de que una medición realizada con este equipo diste de 70 en más de 2 unidades.

```
mean(abs(datos_1$gas_equipo_1-70)>2)
```

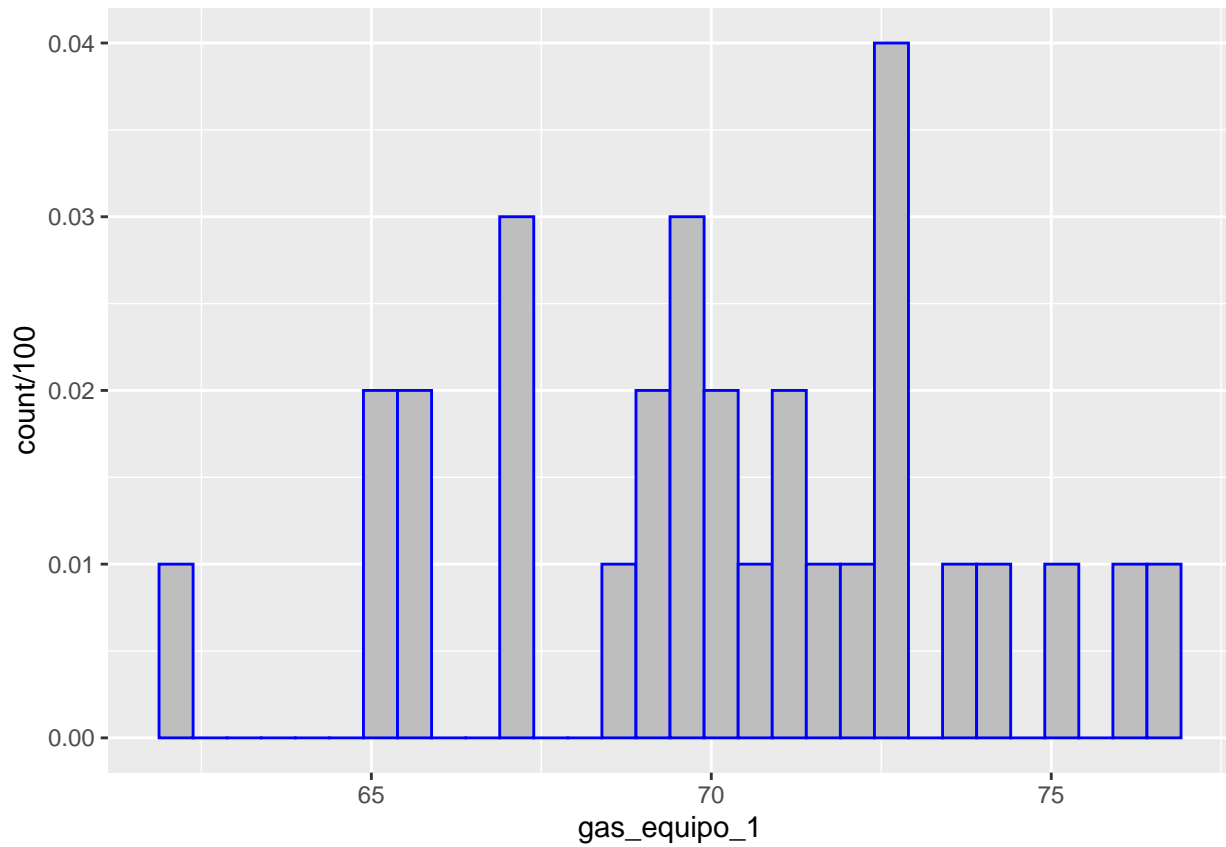
```
## [1] 0.6
```

```
n=30
```

1. Realice un histograma.

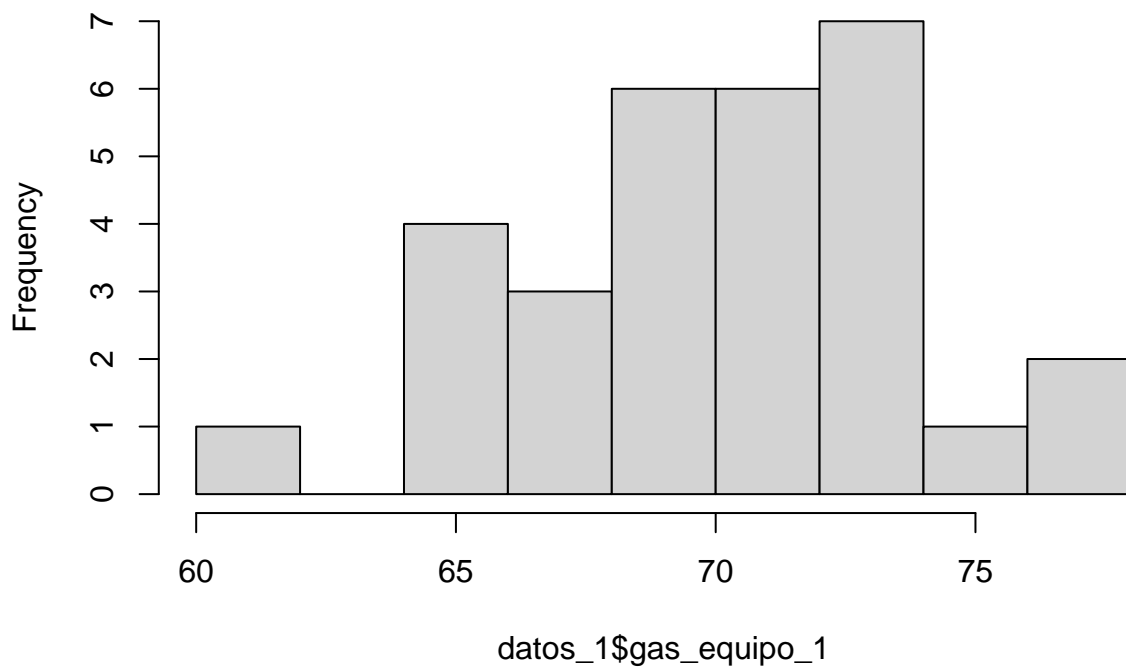
```
n=30
datos_1=read.csv('datos_gas1_n_100.csv', header=TRUE)
datos_1=data.frame('gas_equipo_1'=datos_1[1:n,])
datos_1 %>%
  ggplot()+
  geom_histogram(aes(x=gas_equipo_1,y=stat(count)/100),
    color='blue',fill='grey')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
hist(datos_1$gas_equipo_1)
```

**Histogram of datos\_1\$gas\_equipo\_1**



2. Calcule el promedio de los datos.

```
mean(datos_1$gas_equipo_1)
```

```
## [1] 70.17867
```

3. Calcule el percentil 0.9 de los datos.

```
quantile(datos_1$gas_equipo_1,0.9)
```

```
##      90%
```

```
## 74.041
```

4. Estime la probabilidad de que una medición realizada con este equipo diste de 70 en más de 2 unidades.

```
mean(abs(datos_1$gas_equipo_1-70)>2)
```

```
## [1] 0.6
```

Hagamos ahora la transición de los datos para n en c(5,30,100).

Una forma: nos movemos sobre la cantidad de datos que leemos usando `transition_filter`.

```
datos_1=read.csv('datos_gas1_n_100.csv', header=TRUE)
datos_1$filas=1:dim(datos_1)[1]
# Le agrego una variable estado a los datos
# para agregarle color al grafico.
datos_1$estado=rep(c(1,2,3),c(5,25,70))
anim=datos_1 %>%
  ggplot(aes(fill=estado))+
  geom_histogram(aes(x=gas_equipo_1,y=stat(count)/100))+
  transition_filter(
    transition_length = 3,
    filter_length = 1,
    filas <= 5 ,
    filas <= 30,
    filas <= 100
  )+
  ggtitle(
    'Filter: {closest_filter}',
    subtitle = '{closest_expression}'
  ) +
  enter_fade() +
  exit_fly(y_loc = 0)
# animate(anim,
#           width = 400, height = 400,
#           nframes = 480, fps = 24)
# anim_save("datos 1 transicion color.gif", anim)
```

Otra forma: Nos movemos sobre la variable estado.

```
anim=datos_1 %>%
  ggplot(aes(fill=as.factor(estado)))+
  geom_histogram(aes(x=gas_equipo_1,y=stat(count)/100,group=estado))+
  transition_states(estado, transition_length = 1, state_length = 1)+
  enter_fade() +
  exit_fly(y_loc = 0)
# animate(anim,
#           width = 400, height = 400,
#           nframes = 480, fps = 24)
# anim_save("datos 1 transicion color 2.gif", anim)
```



## Mediciones de gas - Equipo 2

Repita el ejercicio anterior utilizando ahora los datos obtenidos al utilizar el equipo 2.

Resolución:

1. Realizo un histograma

```
datos_2=read.csv('datos_gas2_n_100.csv', header=TRUE)
datos_2$filas=1:dim(datos_2)[1]
datos_2$estado=rep(c(1,2,3),c(5,25,70))
# # creamos una matriz de ( length(enes) ^max(enes) ) ^3
# datos_3d=array(0,rep(max(enes),length(enes)))
# for(i in 1:length(enes)){
#   aux_num=c(1,2,3)
#   datos[,1,i]=data.frame('gas_equipo_2'=datos_2[1:enes(i),])
# }
anim=datos_2 %>%
  ggplot()+
  geom_histogram(aes(x=gas_equipo_2,y=stat(count)/100),
                 color='blue',fill='grey')+
  transition_filter(
    transition_length = 3,
    filter_length = 1,
    filas <= 5 ,
    filas <= 30,
    filas <= 100
  )+
  ggtitle(
    'Filter: {closest_filter}',
    subtitle = '{closest_expression}'
  ) +
  enter_fade() +
  exit_fly(y_loc = 0)
# animate(anim,
#           width = 400, height = 400,
#           nframes = 480, fps = 24)
# anim_save("datos 2 transicion.gif", anim)
```

Otra forma:

```
anim=datos_2 %>%
  ggplot(aes(color=estado))+
  geom_histogram(aes(x=gas_equipo_2,y=stat(count)/100))+
  transition_states(estado, transition_length = 1, state_length = 1)
# animate(anim,
#           width = 400, height = 400,
#           nframes = 480, fps = 24)
# anim_save("datos 2 transicion color.gif", anim)
```

NO ENTIENDO POR QUÉ DICEN QUE ES UNA DISTRIBUCIÓN UNIFORME ¿? PORQUE LA DENSIDAD DE CADA VALOR QUE TOMA EL HISTOGRAMA PARECE ALEATORIO.

2. Calcule el promedio de los datos.

```
mean(datos_2$gas_equipo_2)
```

```
## [1] 70.2186
```

3. Calcule el percentil 0.9 de los datos.

```
quantile(datos_2$gas_equipo_2,0.9)
```

```
##      90%
```

```
## 72.402
```

4. Estime la probabilidad de que una medición realizada con este equipo diste de 70 en más de 2 unidades.

```
mean(abs(datos_2$gas_equipo_2-70)>2)
```

```
## [1] 0.37
```

## Duración de lámparas

Considere los datos de la duración de  $n$  lámparas (en meses), para  $n$  en  $\{5, 30, 100\}$ . En cada caso,

1. realice un histograma.
2. calcule el promedio y el percentil 0.9.
3. estime la probabilidad de que la lámpara dure a lo sumo un año (12 meses).

Resolución:

1. Realice un histograma.

```
lamparas=read.csv('datos_lamparas_n_100.csv', header=TRUE)
lamparas$estado=rep(c(1,2,3),c(5,25,70))
anim=lamparas %>%
  ggplot(aes(color=estado))+
  geom_histogram(aes(x=lamparas,y=stat(count)/100))+
  transition_states(estado, transition_length = 1, state_length = 1)
# animate(anim,
#           width = 400, height = 400,
#           nframes = 480, fps = 24)
# anim_save("datos lamparas transicion.gif", anim)
```

2. calcule el promedio y el percentil 0.9.

```
mean(lamparas$lamparas)
```

```
## [1] 10.046
```

```
quantile(lamparas$lamparas,0.9)
```

```
## 90%
```

```
## 25.53
```

3. estime la probabilidad de que la lámpara dure a lo sumo un año (12 meses).

```
mean(lamparas$lamparas<12)
```

```
## [1] 0.68
```

### 3. Parte 2 - El mono y la mesa

1. Proponer un modelo para los datos generados por el mono e identificar el parámetro de interés.
2. Proponer estimadores para el parámetro de interés.
3. Calcular los estimadores con los datos que le hemos asignado en

Resolución:

1. Proponemos el modelo  $\mathcal{U}(0, \theta)$  con  $\theta$  = longitud de la mesa porque el mono saca valores al azar y a lo sumo pueden valer el largo de la mesa.
2. Proponemos los estimadores:
  - $2 * \overline{X_n}$ : motivados en que  $\overline{X_n} \xrightarrow{n \rightarrow \infty} \frac{\theta}{2}$ ;
  - $\max(X_1, \dots, X_n)$ : motivados en que  $\theta$  es el largo de la mesa, entonces la máxima longitud que genere el mono va a ser lo mas parecido a  $\theta$ .
3. Calculo dos estimadores de los 5 datos del documento compartido (hoja=Datos mesa).

```
datos=c(1.7, 2.48, 0.97, 2.22, 1.83)
est1=function(datos){
  2*mean(datos)
}
est1(datos)
```

```
## [1] 3.68
```

```
est2=function(datos){
  max(datos)
}
est2(datos)
```

```
## [1] 2.48
```