

Visualizacion-tarea

November 13, 2020

1 Visualización Tarea

```
[1]: library(ggplot2)
library(tidyverse)
library(ggthemes)
library(plotly)
```

```
Attaching packages
1.3.0

tidyverse

tibble 3.0.3      dplyr  1.0.1
tidyr  1.1.1      stringr 1.4.0
readr  1.3.1      forcats 0.5.0
purrr  0.3.4
```

```
Conflicts
tidyverse_conflicts()
dplyr::filter() masks stats::filter()
dplyr::lag()    masks stats::lag()
```

Attaching package: ‘plotly’

The following object is masked from ‘package:ggplot2’:

last_plot

The following object is masked from ‘package:stats’:

filter

The following object is masked from ‘package:graphics’:

layout

2 Parte 1

De la base de datos de proyecciones poblacionales de CONAPO ([aquí](#) la base), hacer al menos 5 visualizaciones para entender el crecimiento poblacional. Se tienen muchas variables y categorías por explorar, se espera un análisis y contexto de las gráficas así como una interpretación.

```
[2]: CONAPO <- read.csv(file="pob_mit_proyecciones.csv", fileEncoding = 'iso-8859-1')
      head(CONAPO)
```

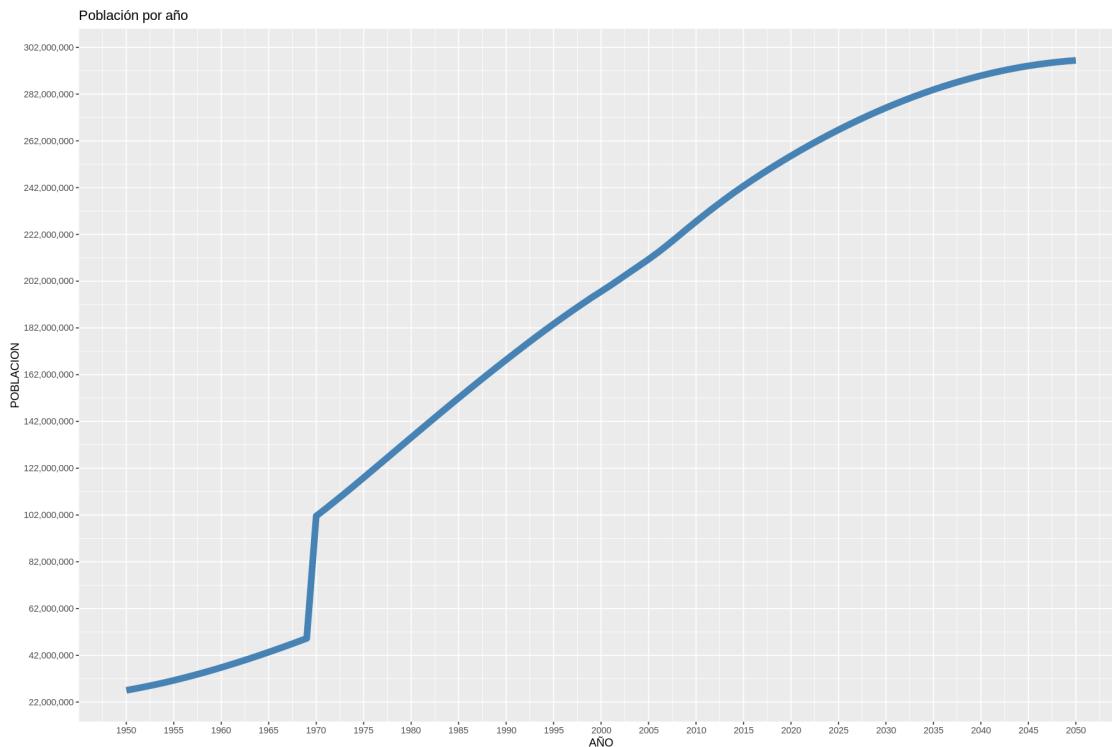
		REGLON	AÑO	ENTIDAD	CVE_GEO	EDAD	SEXO	POBLA
		<int>	<int>	<fct>	<int>	<int>	<fct>	<int>
A data.frame: 6 × 7	1	1	1950	República Mexicana	0	0	Hombres	572103
	2	2	1950	República Mexicana	0	0	Mujeres	559162
	3	3	1950	República Mexicana	0	1	Hombres	514540
	4	4	1950	República Mexicana	0	1	Mujeres	505269
	5	5	1950	República Mexicana	0	2	Hombres	478546
	6	6	1950	República Mexicana	0	2	Mujeres	469322

```
[3]: options(repr.plot.width=15, repr.plot.height=10)
      CONAPO %>%
        group_by(AÑO) %>%
        summarise(POBLACION = sum(POBLACION)) %>%
        ggplot(aes(x=AÑO,y=POBLACION)) +
          geom_line(colour="steelblue", size = 3) +
          labs(title = "Población por año") +
          ylim(2000000,300000000) +
          xlim(1950, 2050) +
          scale_x_continuous(breaks=seq(from = 1950, to = 2050, by = 5)) +
          scale_y_continuous(breaks=seq(from = 2000000, to = 400000000, by = 20000000), labels = scales::comma)
```

`summarise()` ungrouping output (override with `.groups` argument)

Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing scale.

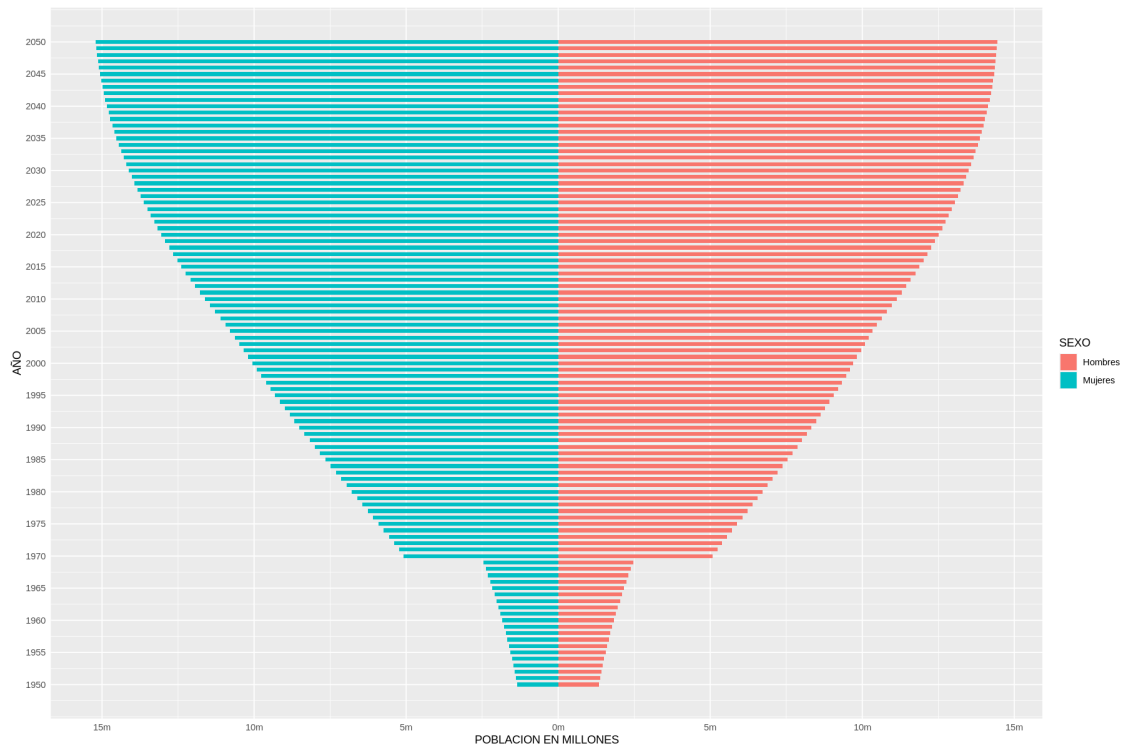
Scale for 'y' is already present. Adding another scale for 'y', which will replace the existing scale.



- Se puede observar que alrededor de 1970 hubo un gran incremento en la población, esto puede ser porque se comenzó a contabilizar más formalmente a la población.
- Entre 1970 y 2015 se puede observar un crecimiento casi constante de la población.
- A partir del año 2015 y con los pronósticos hacia el 2050 se espera que la población llegue a estabilizarse probablemente porque está llegando al nivel máximo de población M.

```
[5]: CONAPO %>%
  group_by(SEXO, AÑO) %>%
  summarise(POBLACION = sum(POBLACION)) %>%
  ggplot(aes(x=AÑO, y=ifelse(SEXO=='Mujeres', -POBLACION, POBLACION), fill =
  ↪ SEXO)) +
  geom_bar(stat = "identity", width = .6)+
  scale_x_continuous(breaks=seq(from = 1950, to = 2050, by = 5)) +
  scale_y_continuous(breaks=seq(from = -150000000, to = 150000000, by =
  ↪ 50000000),
                    labels = paste0(as.character(c(seq(15, 0, -5), seq(5,
  ↪ 15, 5))), "m")) +
  ylab("POBLACION EN MILLONES")+
  coord_flip() +
  theme(plot.title = element_text(hjust = .5), axis.ticks = element_blank())
```

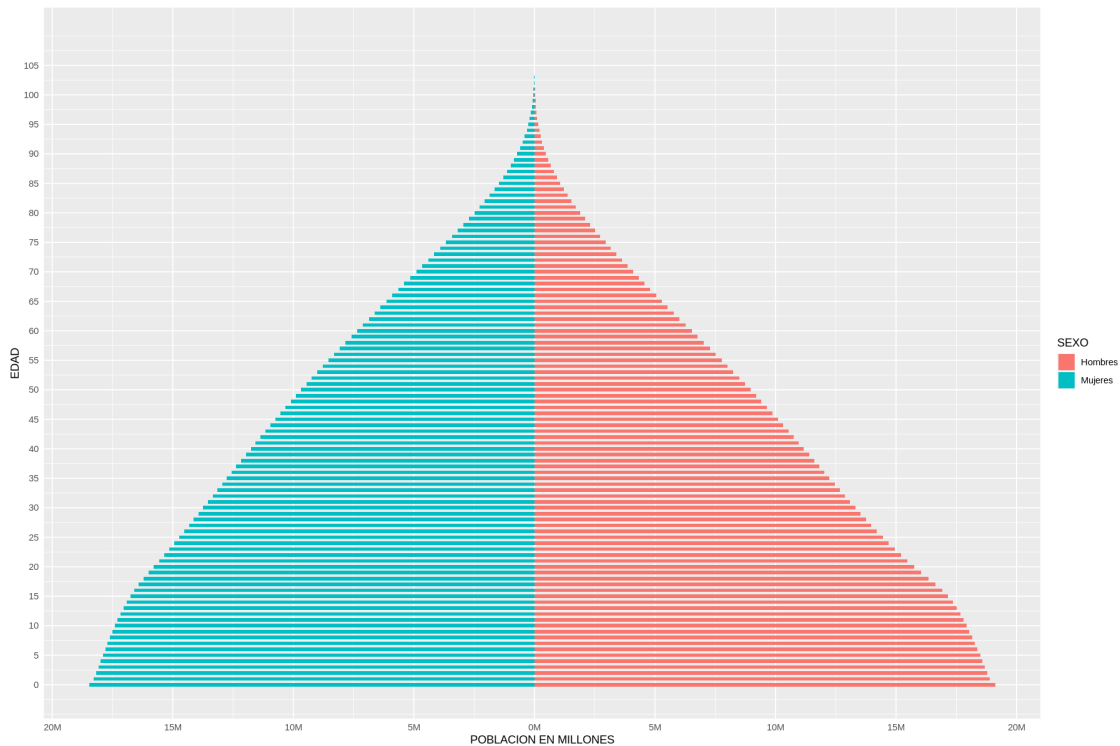
`summarise()` regrouping output by 'SEXO' (override with `groups` argument)



- Podemos observar que a lo largo de los años prácticamente se ha mantenido equivalente la proporción entre hombres y mujeres.
- Después del año 2000 se puede apreciar una mayor proporción de mujeres que de hombres

```
[6]: CONAPO %>%
  group_by(SEXO, EDAD) %>%
  summarise(POBLACION = sum(POBLACION)) %>%
  ggplot(aes(x=EDAD,y=ifelse(SEXO=='Mujeres', -POBLACION, POBLACION), fill =
  ↪SEXO)) +
  geom_bar(stat = "identity", width = .6)+
  scale_x_continuous(breaks=seq(from = 0, to = 105, by = 5)) +
  scale_y_continuous(breaks=seq(from = -200000000, to = 200000000, by =
  ↪50000000),
                    labels = paste0(as.character(c(seq(20, 0, -5), seq(5,
  ↪20, 5))), "M")) +
  ylab("POBLACION EN MILLONES")+
  coord_flip() +
  theme(plot.title = element_text(hjust = .5), axis.ticks = element_blank())
```

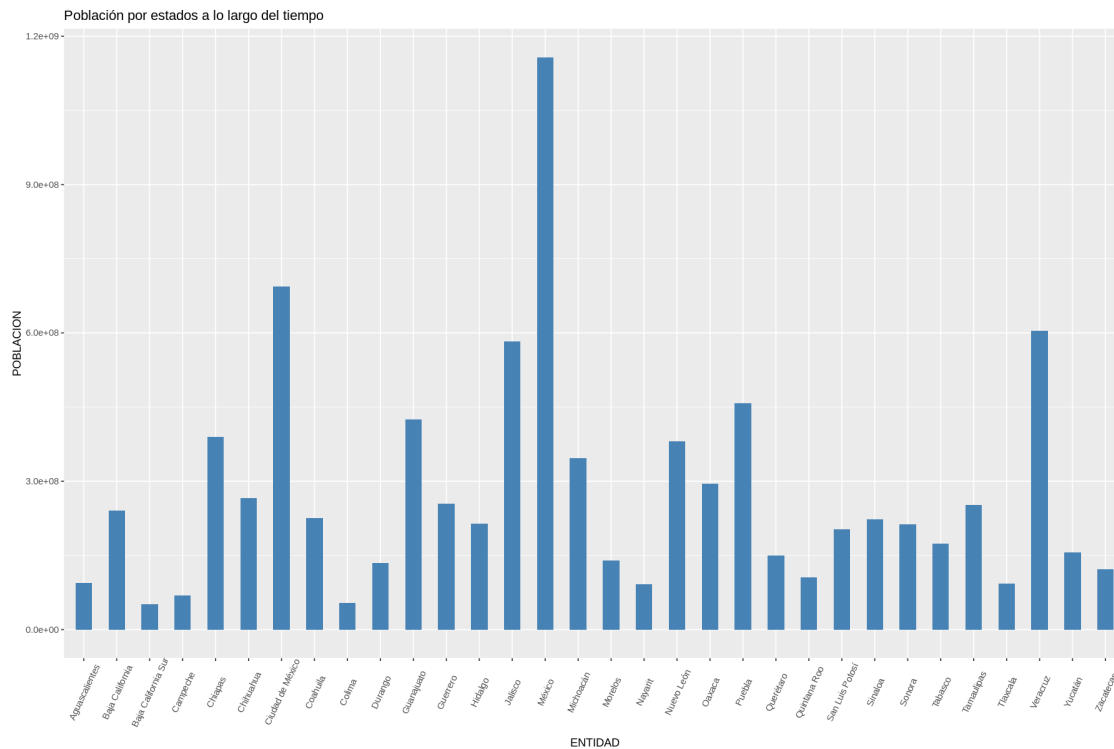
`summarise()` regrouping output by 'SEXO' (override with `groups` argument)



- Entre los 15 y 20 años hay una mayor proporción de hombres que de mujeres
- La esperanza de vida de las mujeres es mayor que la de los hombres

```
[7]: CONAPO %>%
  group_by(ENTIDAD) %>%
  summarise(POBLACION = sum(POBLACION)) %>%
  subset(ENTIDAD!="República Mexicana") %>%
  ggplot(aes(ENTIDAD, POBLACION)) +
    geom_bar(stat="identity", width = 0.5, fill="steelblue") +
    labs(title="Población por estados a lo largo del tiempo") +
    theme(axis.text.x = element_text(angle=65, vjust=0.6))
```

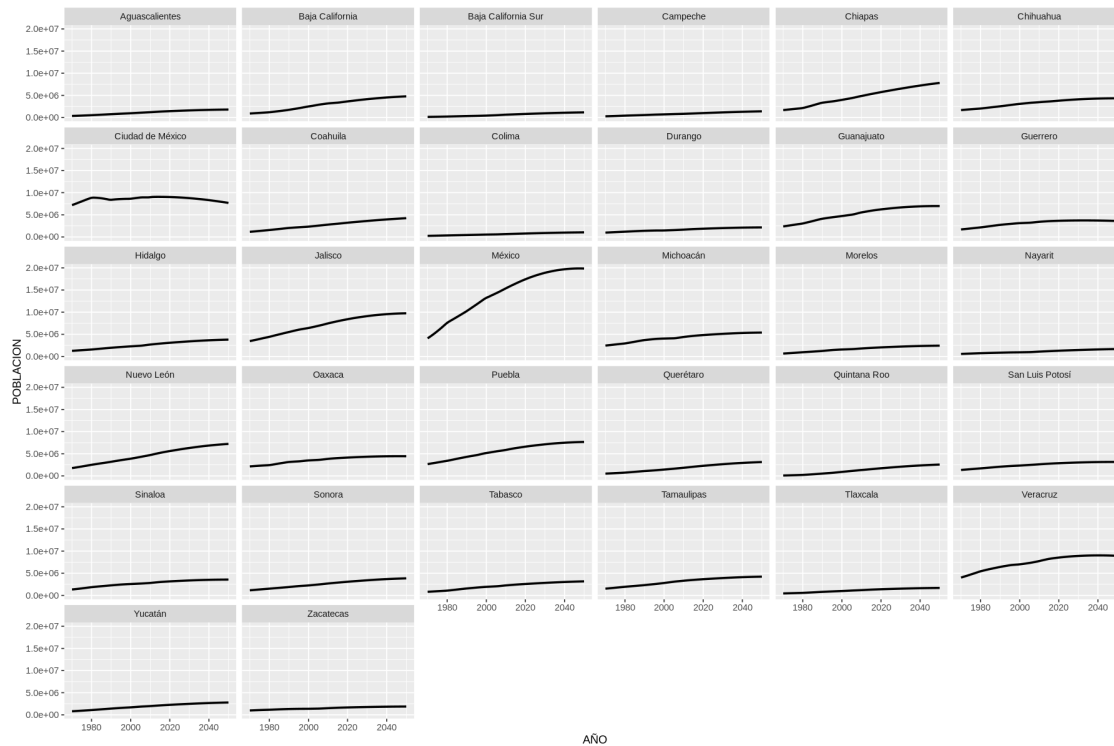
`summarise()` ungrouping output (override with `.groups` argument)



- Estado de México, CDMX y Veracruz son los estados más poblados.
- Baja California Sur, Colima y Campeche son los estados menos poblados

```
[8]: CONAPO %>%
      group_by(ENTIDAD, AÑO) %>%
      summarise(POBLACION = sum(POBLACION)) %>%
      subset(ENTIDAD!="República Mexicana") %>%
      ggplot(aes(AÑO, POBLACION)) +
        geom_line(size=1) +
        facet_wrap(~ENTIDAD)
```

```
`summarise()` regrouping output by 'ENTIDAD' (override with `.groups` argument)
```



- Zacatecas ha aumentado su población muy lentamente
- Estado de México ha aumentado su población muy rápidamente y está llegando a su límite
- Se espera que la CDMX comience a reducir su población en los próximos años

3 Parte 2

Con la misma base de datos, se requiere hacer todas las visualizaciones interactivas; las mismas 5 o más visualizaciones, ahora interactivas (recuerden plot.ly). Sugerencia: el año tiene que ser una de estas categorías ya que son proyecciones.

```
[10]: a <-CONAPO %>%
  group_by(AÑO) %>%
  summarise(POBLACION = sum(POBLACION)) %>%
  ggplot(aes(x=AÑO,y=POBLACION)) +
    geom_line(colour="steelblue", size = 3) +
    labs(title = "Población por año") +
    ylim(2000000,300000000) +
    xlim(1950, 2050) +
    scale_x_continuous(breaks=seq(from = 1950, to = 2050, by = 5)) +
    scale_y_continuous(breaks=seq(from = 2000000, to = 400000000, by = 20000000))
  ggplotly(a)
```

`summarise()` ungrouping output (override with `.groups` argument)

Scale for 'x' is already present. Adding another scale for 'x', which will replace the existing scale.

Scale for 'y' is already present. Adding another scale for 'y', which will replace the existing scale.

HTML widgets cannot be represented in plain text (need html)

```
[11]: b <-CONAPO %>%
      group_by(SEXO, AÑO) %>%
      summarise(POBLACION = sum(POBLACION)) %>%
      ggplot(aes(x=AÑO,y=ifelse(SEXO=='Mujeres', -POBLACION, POBLACION), fill=
      ↪ SEXO)) +
      geom_bar(stat = "identity", width = .6)+
      scale_x_continuous(breaks=seq(from = 1950, to = 2050, by = 5)) +
      scale_y_continuous(breaks=seq(from = -150000000, to = 150000000, by =
      ↪ 50000000),
                        labels = paste0(as.character(c(seq(15, 0, -5),
      ↪ seq(5, 15, 5))), "m")) +
      ylab("POBLACION EN MILLONES")+
      coord_flip() +
      theme(plot.title = element_text(hjust = .5), axis.ticks =
      ↪ element_blank())

      ggplotly(b)
```

`summarise()` regrouping output by 'SEXO' (override with `.groups` argument)

HTML widgets cannot be represented in plain text (need html)

```
[12]: c <-CONAPO %>%
      group_by(SEXO, EDAD) %>%
      summarise(POBLACION = sum(POBLACION)) %>%
      ggplot(aes(x=EDAD,y=ifelse(SEXO=='Mujeres', -POBLACION, POBLACION),
      ↪ fill = SEXO)) +
      geom_bar(stat = "identity", width = .6)+
      scale_x_continuous(breaks=seq(from = 0, to = 105, by = 5)) +
      scale_y_continuous(breaks=seq(from = -200000000, to = 200000000, by =
      ↪ 50000000),
                        labels = paste0(as.character(c(seq(20, 0, -5),
      ↪ seq(5, 20, 5))), "M")) +
      ylab("POBLACION EN MILLONES")+
      coord_flip() +
```



```

    theme(plot.title = element_text(hjust = .5), axis.ticks =
↪element_blank())

ggplotly(c)

```

`summarise()` regrouping output by 'SEXO' (override with `.groups` argument)

HTML widgets cannot be represented in plain text (need html)

```

[13]: d <-CONAPO %>%
      group_by(ENTIDAD) %>%
      summarise(POBLACION = sum(POBLACION)) %>%
      subset(ENTIDAD!="República Mexicana") %>%
      ggplot(aes(ENTIDAD, POBLACION)) +
        geom_bar(stat="identity", width = 0.5, fill="steelblue") +
        labs(title="Población por estados a lo largo del tiempo") +
        theme(axis.text.x = element_text(angle=65, vjust=0.6))

ggplotly(d)

```

`summarise()` ungrouping output (override with `.groups` argument)

HTML widgets cannot be represented in plain text (need html)

```

[14]: e <-CONAPO %>%
      group_by(ENTIDAD, AÑO) %>%
      summarise(POBLACION = sum(POBLACION)) %>%
      subset(ENTIDAD!="República Mexicana") %>%
      ggplot(aes(AÑO, POBLACION)) +
        geom_line(size=1) +
        facet_wrap(~ENTIDAD)

ggplotly(e)

```

`summarise()` regrouping output by 'ENTIDAD' (override with `.groups` argument)

Warning message:

"`group_by()` is deprecated as of dplyr 0.7.0.

Please use `group_by()` instead.

See vignette('programming') for more help

This warning is displayed once every 8 hours.

Call `lifecycle::last_warnings()` to see where this warning was

generated."

HTML widgets cannot be represented in plain text (need html)

4 Parte 3

Hacer 1 ejercicio de cada sección

4.1 Ejercicios intro

4.1.1 Velocidad de la luz

Tenemos 5 experimentos con 20 corridas de mediciones de la velocidad de la luz.

- ¿Cómo exponen los resultados?
- ¿Los resultados se ven equivalentes entre los experimentos?

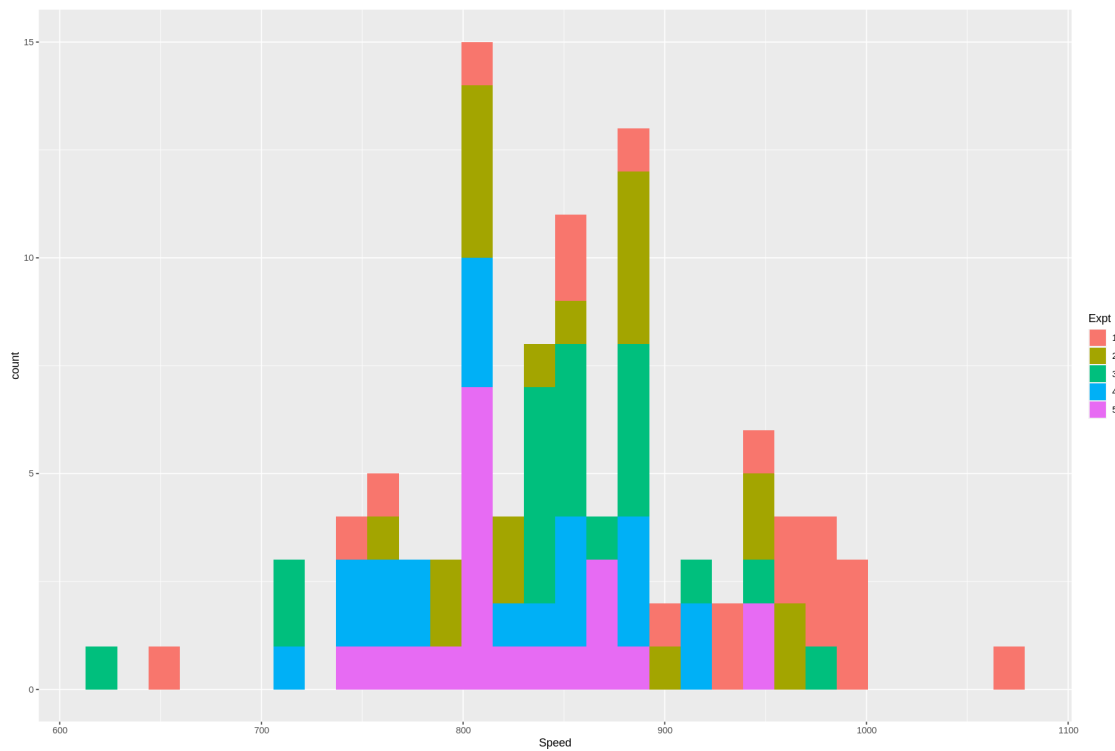
```
[16]: library(MASS)
```

```
[18]: vl <- data.frame(michelson)
      vl
```

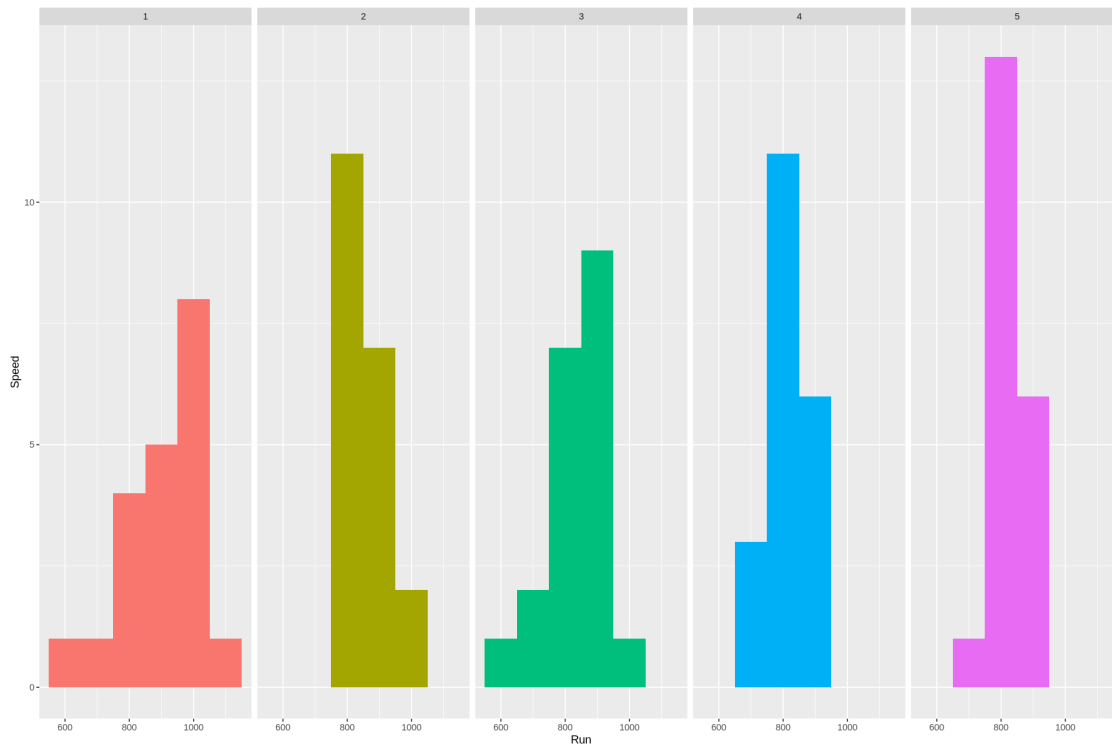
	Speed <int>	Run <fct>	Expt <fct>
1	850	1	1
2	740	2	1
3	900	3	1
4	1070	4	1
5	930	5	1
6	850	6	1
7	950	7	1
8	980	8	1
9	980	9	1
10	880	10	1
11	1000	11	1
12	980	12	1
13	930	13	1
14	650	14	1
15	760	15	1
16	810	16	1
17	1000	17	1
18	1000	18	1
19	960	19	1
20	960	20	1
21	960	1	2
22	940	2	2
23	960	3	2
24	940	4	2
25	880	5	2
26	800	6	2
27	850	7	2
28	880	8	2
29	900	9	2
30	840	10	2
71	910	11	4
72	920	12	4
73	890	13	4
74	860	14	4
75	880	15	4
76	720	16	4
77	840	17	4
78	850	18	4
79	850	19	4
80	780	20	4
81	890	1	5
82	840	2	5
83	780	3	5
84	810	4	5
85	760	5	5
86	810	6	5
87	790	7	5
88	810	8	5
89	820	9	5
90	850	10	5

```
[40]: ggplot(v1, aes(Speed, fill=Expt)) +  
       geom_histogram()
```

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



```
[41]: ggplot(v1, aes(Speed, fill=Expt)) +  
       geom_histogram(binwidth = 100) +  
       xlab("Run") +  
       ylab("Speed") +  
       facet_grid(~Expt) +  
       theme(legend.position = "none")
```

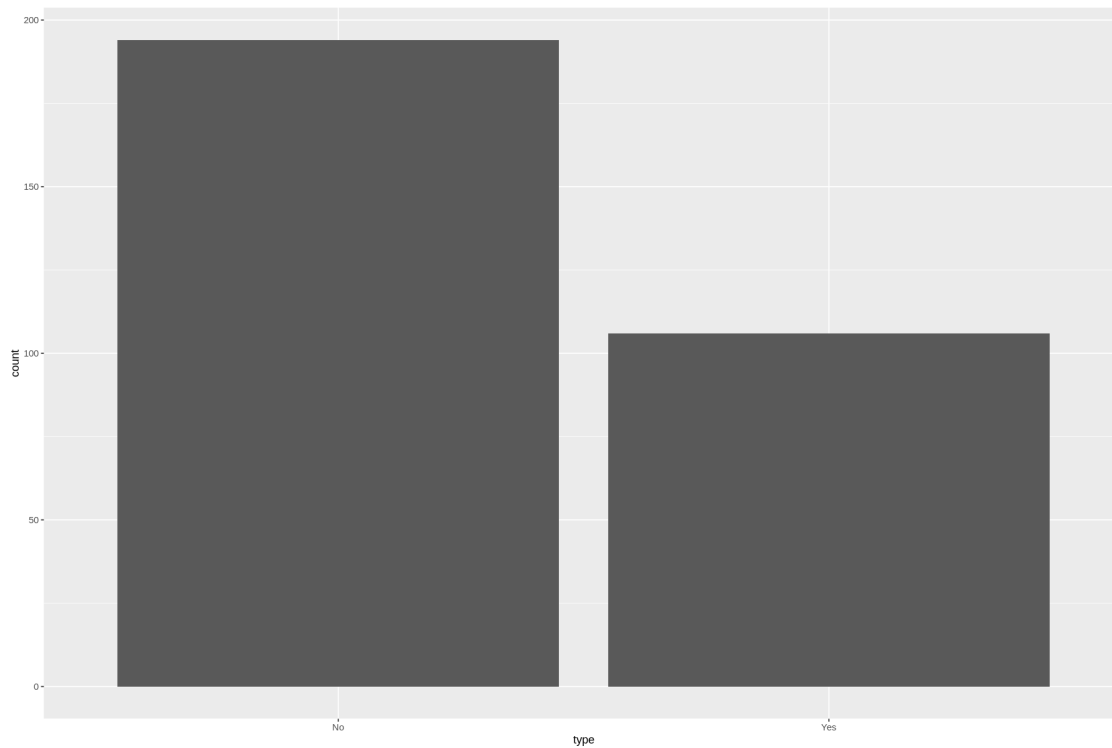


Podemos observar que los resultados se ven uniformes entre los experimentos y que la mayor concentración se encuentra entre 800 y 900

4.1.2 Pima Indians

- ¿Qué dice este gráfico de barras?

```
[66]: ggplot(Pima.tr2, aes(type))+geom_bar()
```



Son las personas que tienen diabetes.

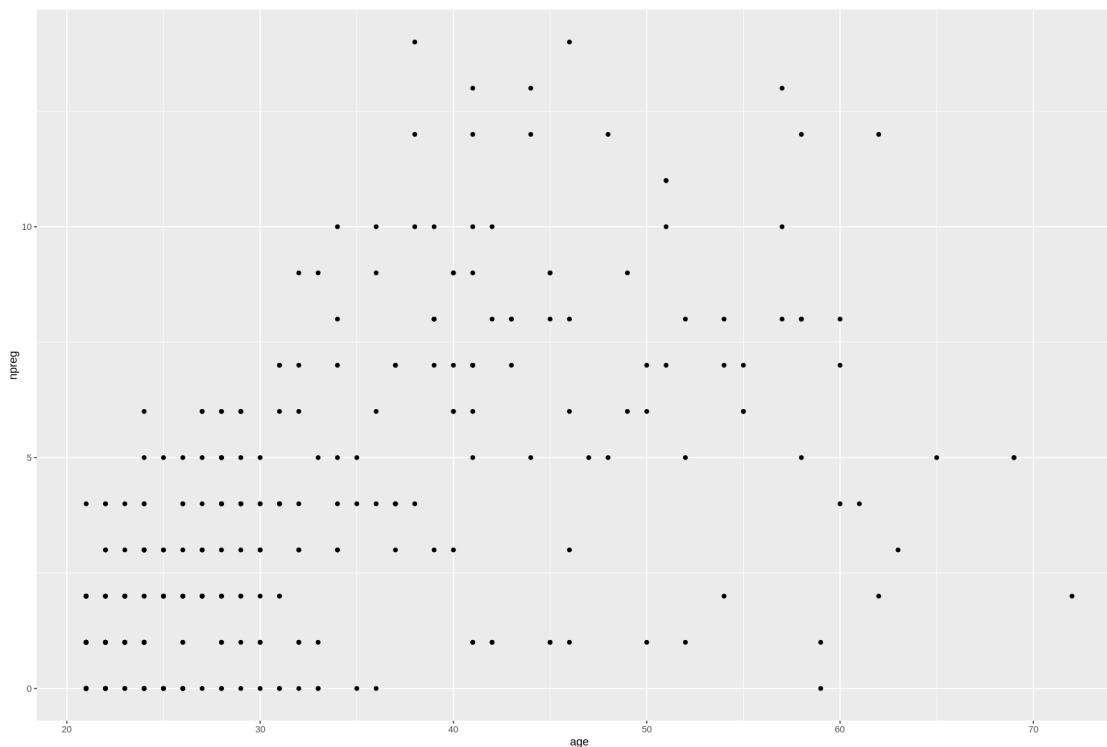
[65]: `Pima.tr2`

	npreg	glu	bp	skin	bmi	ped	age	type	
	<int>	<int>	<int>	<int>	<dbl>	<dbl>	<int>	<fct>	
A data.frame: 300 × 8	1	5	86	68	28	30.2	0.364	24	No
	2	7	195	70	33	25.1	0.163	55	Yes
	3	5	77	82	41	35.8	0.156	35	No
	4	0	165	76	43	47.9	0.259	26	No
	5	0	107	60	25	26.4	0.133	23	No
	6	5	97	76	27	35.6	0.378	52	Yes
	7	3	83	58	31	34.3	0.336	25	No
	8	1	193	50	16	25.9	0.655	24	No
	9	3	142	80	15	32.4	0.200	63	No
	10	2	128	78	37	43.3	1.224	31	Yes
	11	0	137	40	35	43.1	2.288	33	Yes
	12	9	154	78	30	30.9	0.164	45	No
	13	1	189	60	23	30.1	0.398	59	Yes
	14	12	92	62	7	27.6	0.926	44	Yes
	15	1	86	66	52	41.3	0.917	29	No
	16	4	99	76	15	23.2	0.223	21	No
	17	1	109	60	8	25.4	0.947	21	No
	18	11	143	94	33	36.6	0.254	51	Yes
	19	1	149	68	29	29.3	0.349	42	Yes
	20	0	139	62	17	22.1	0.207	21	No
	21	2	99	70	16	20.4	0.235	27	No
	22	1	100	66	29	32.0	0.444	42	No
	23	4	83	86	19	29.3	0.317	34	No
	24	0	101	64	17	21.0	0.252	21	No
	25	1	87	68	34	37.6	0.401	24	No
	26	9	164	84	21	30.8	0.831	32	Yes
	27	1	99	58	10	25.4	0.551	21	No
	28	0	140	65	26	42.6	0.431	24	Yes
	29	5	108	72	43	36.1	0.263	33	No
	30	2	110	74	29	32.4	0.698	27	No
	271	2	87	NA	23	28.9	0.773	25	No
	272	10	108	66	NA	32.4	0.272	42	Yes
	273	10	139	80	NA	27.1	1.441	57	No
	274	4	110	92	NA	37.6	0.191	30	No
	275	4	114	64	NA	28.9	0.126	24	No
	276	0	101	62	NA	21.9	0.336	25	No
	277	2	91	62	NA	27.3	0.525	22	No
	278	8	133	72	NA	32.9	0.270	39	Yes
	279	1	111	94	NA	32.8	0.265	45	No
	280	5	147	75	NA	29.9	0.434	28	No
	281	4	92	80	NA	42.2	0.237	29	No
	282	2	90	60	NA	23.5	0.191	25	No
	283	7	114	64	NA	27.4	0.732	34	Yes
	284	7	125	86	NA	37.6	0.304	51	No
	285	2	119	NA	NA	19.6	0.832	72	No
	286	5	115	76	NA	31.2	0.343	44	Yes
	287	0	141	NA	NA	42.4	0.205	29	Yes
	288	0	167	NA	NA	32.3	0.839	30	Yes
	289	4	90	NA	NA	28.0	0.610	31	No
	290	5	132	80	NA	26.8	0.186	69	No

```
[ ]: help(Pima.tr2)
```

- ¿Por qué la parte superior izquierda está vacía?

```
[67]: ggplot(Pima.tr2, aes(age,npreg))+geom_point()
```



No hay muchas mujeres con muchos embarazos a los 20 años

4.1.3 Titanic

```
[62]: head(data.frame(Titanic))
```

A data.frame: 6 × 5

	Class	Sex	Age	Survived	Freq
	<fct>	<fct>	<fct>	<fct>	<dbl>
1	1st	Male	Child	No	0
2	2nd	Male	Child	No	0
3	3rd	Male	Child	No	35
4	Crew	Male	Child	No	0
5	1st	Female	Child	No	0
6	2nd	Female	Child	No	0

```
[63]: data.frame(Titanic) %>%
  group_by(Class) %>%
  summarise(count = sum(Freq))
```



```
`summarise()` ungrouping output (override with `.groups` argument)
```

```

      Class    count
    <fct>   <dbl>
1st       325
2nd       285
3rd       706
Crew      885

```

A tibble: 4 × 2

- ¿Qué se les ocurre para graficar todo?
- ¿Cuántos pasajeros en total viajaban por clase?
- Exploren una por una las otras variables categóricas del *dataset*

4.1.4 Swiss

Tenemos datos de fertilidad y otras variables socioeconómicas de 47 provincias franco-parlantes en Suiza en 1888.

```
[68]: head(data.frame(swiss))
```

		Fertility <dbl>	Agriculture <dbl>	Examination <int>	Education <int>	Catholic <dbl>	Infant.Morta <dbl>
	Courtelary	80.2	17.0	15	12	9.96	22.2
	Delemont	83.1	45.1	6	9	84.84	22.2
	Franches-Mnt	92.5	39.7	5	5	93.40	20.2
	Moutier	85.8	36.5	12	7	33.77	20.3
	Neuveville	76.9	43.5	17	15	5.16	20.6
	Porrentruy	76.1	35.3	9	7	90.57	26.6

A data.frame: 6 × 6

- ¿Cómo conjuntarían todas las variables?
- Hagan gráficas para cada variable. ¿Se ve algo especial o raro?
- Hagan un scatterplot de Fertilidad vs el porcentaje de católicos. ¿Se ve algo interesante?
- ¿Hay relación entre agricultura y Educación?

4.2 Ejercicios variables continuas

4.2.1 Galaxias

Usando el conjunto *galaxies* de MASS, que contiene velocidades para 82 planetas.

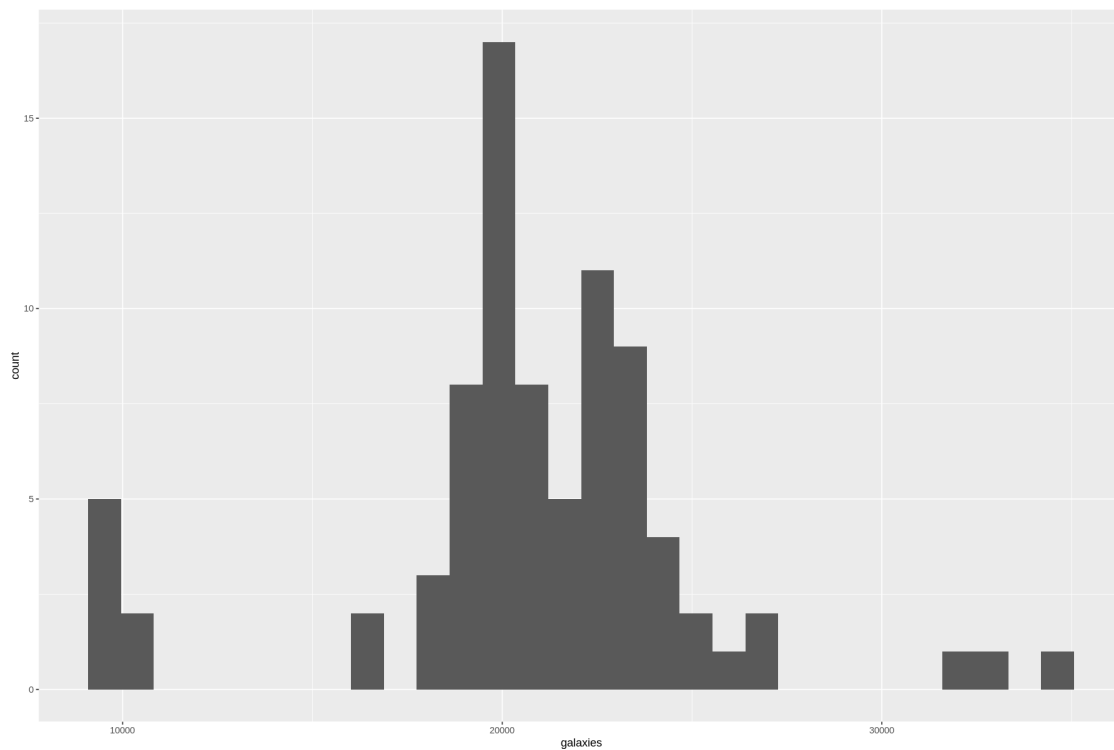
1. Hagan histogramas, boxplots y la aproximación de la densidad.
2. Cambien los anchos de las bandas y expliquen cuál es el mejor?
3. Elijan el número de gráficas que se requieren para este experimento

```
[89]: data(galaxies, package="MASS")
Galaxies <- data_frame(galaxies)
head(Galaxies)
```

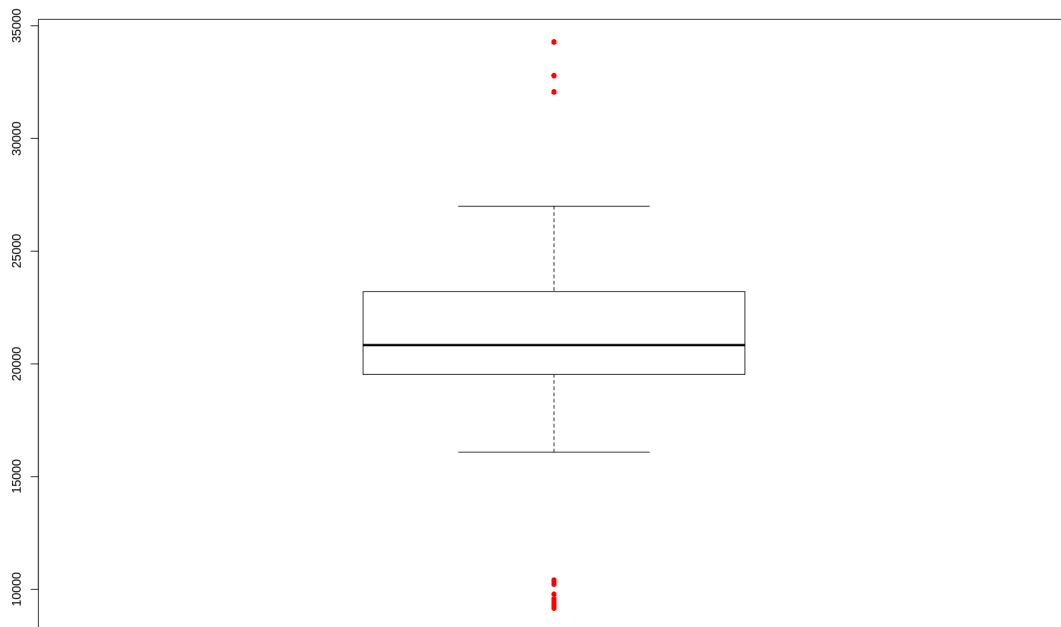
	galaxies
	<dbl>
A tibble: 6 × 1	9172
	9350
	9483
	9558
	9775
	10227

```
[92]: Galaxies %>%
  ggplot(aes(galaxies)) +
  geom_histogram()
```

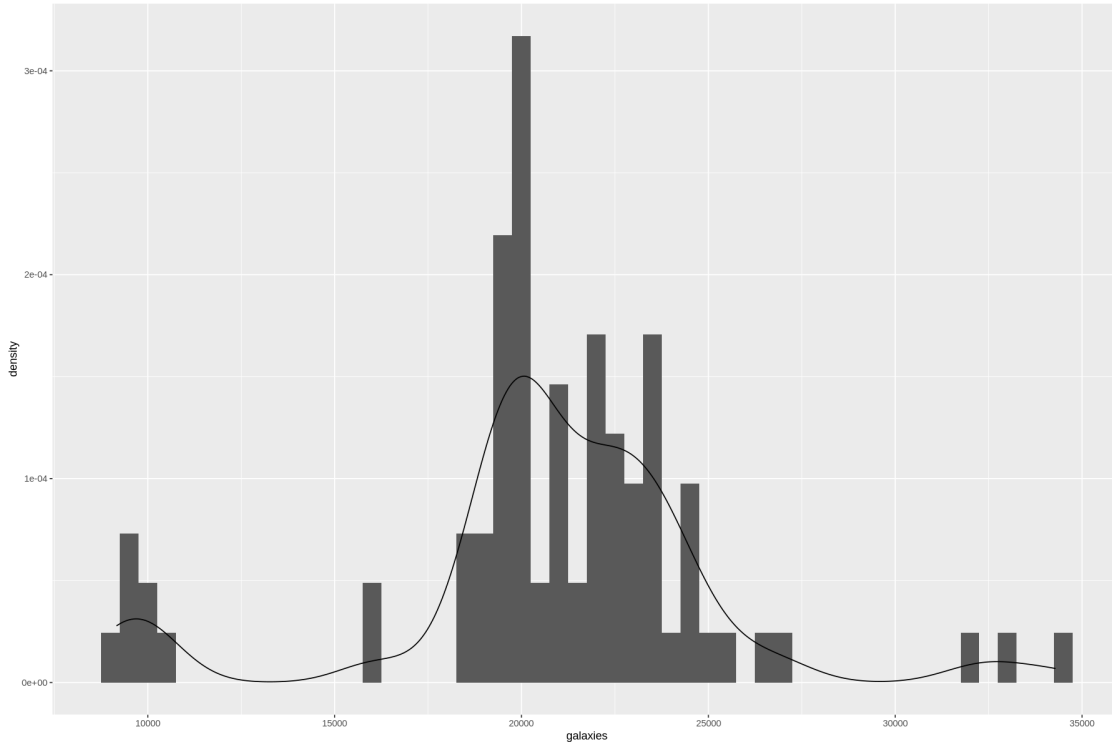
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



```
[94]: Galaxies %>%
  boxplot(pch=16, outcol="red")
```



```
[107]: Galaxies %>%  
  ggplot(aes(galaxies)) +  
    geom_histogram(aes(y = ..density..), binwidth=500) +  
    geom_density()
```



El mejor ancho de banda fue aquel que permitió visualizar los datos de manera más clara.

4.2.2 Estudiantes.

El conjunto *survey* contiene información de estudiantes tomando su primer curso de estadística:

1. Hacer el histograma y poner encima la estimación de la densidad, ¿hay bimodalidad?
2. Jueguen con los anchos para tener mejores estimaciones de la densidad, ¿cuál es mejor?
3. Comparen las distribuciones de hombres y mujeres, que compartan la escala, con distintas estimaciones de densidad.

4.2.3 Presupuesto

El conjunto *zuni* del paquete *lawstat*, contiene 3 variables, distrito, ingreso por estudiante en dólares y el número de estudiantes.

1. ¿Considerarías el 5% más bajo outliers o extremos?
2. Quitando el 5% inferior hagan el gráfico de estimación de densidad ¿Es simétrico?
3. Hagan un Q-Q plot y comenten si es normal o no.

4.3 Ejercicios variables categóricas

4.3.1 Máquinas tragamonedas

En el paquete DAAG, está el dataset *vlt*, muestren si los símbolos que aparecen tienen la misma frecuencia, o no. Ejes comparables.

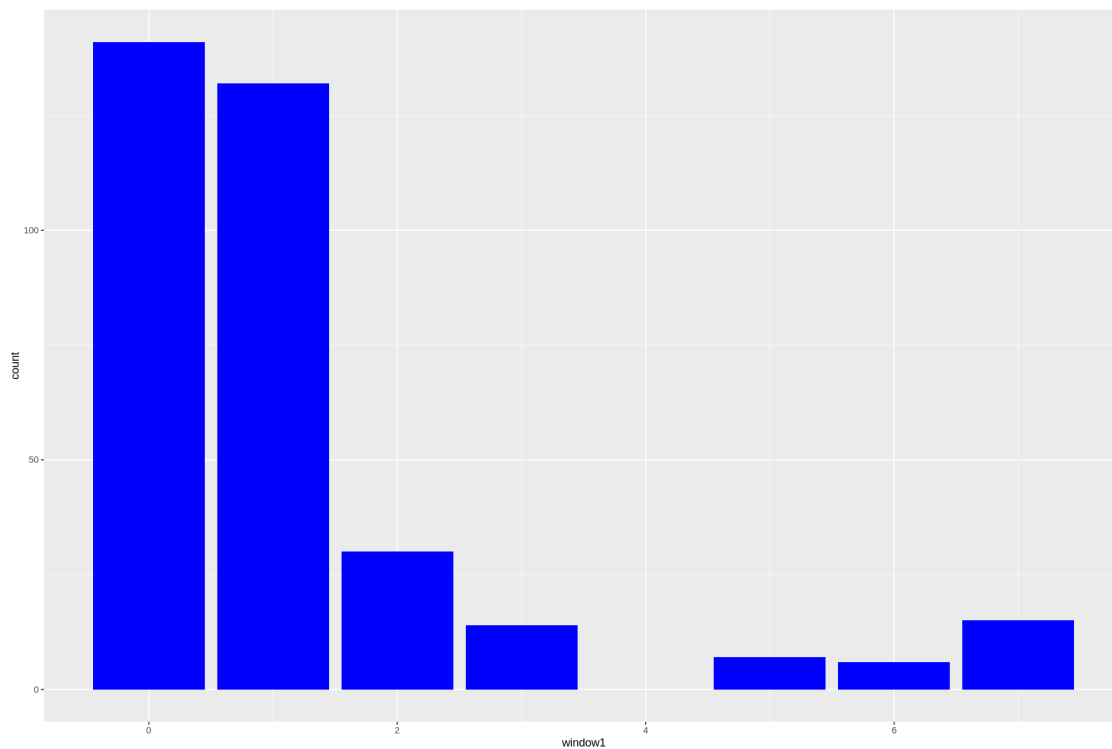
```
[42]: data(vlt, package="DAAG")
      head(vlt)
```

A data.frame: 6 × 5

	window1 <int>	window2 <int>	window3 <int>	prize <int>	night <int>
1	2	0	0	0	1
2	0	5	1	0	1
3	0	0	0	0	1
4	2	0	0	0	1
5	0	0	0	0	1
6	0	0	1	0	1

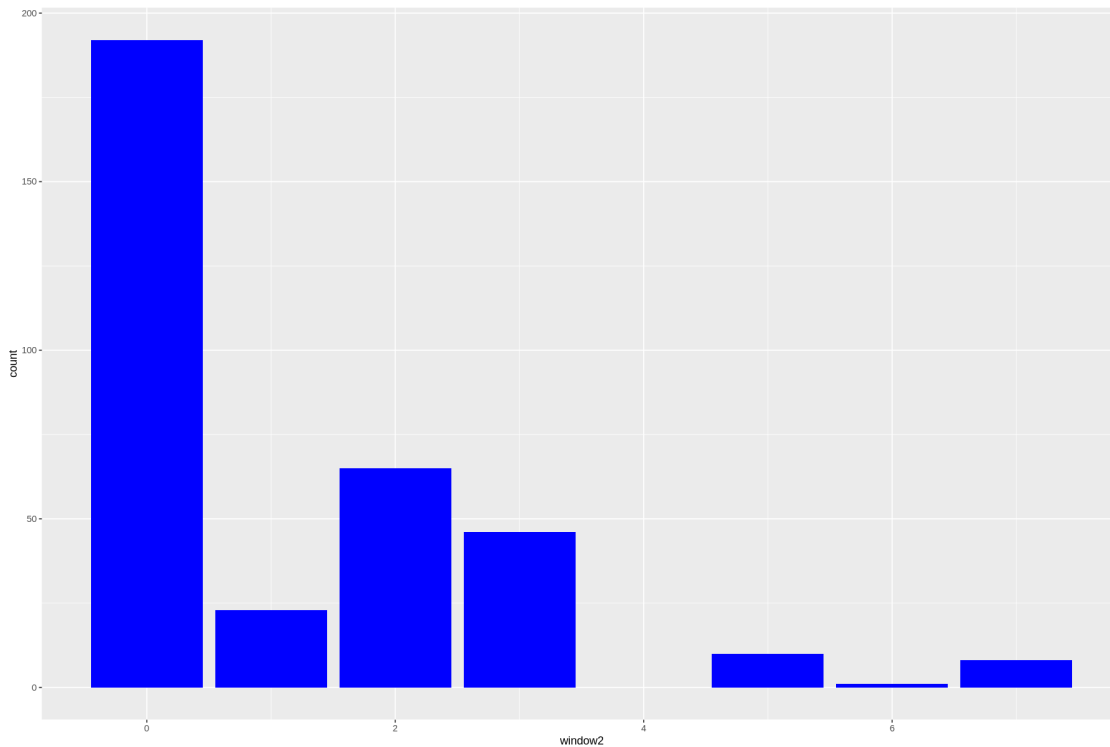
```
[43]: vlt %>%
      group_by(window1) %>%
      summarise(n = n()) %>%
      ggplot(aes(weight=n)) +
        aes(window1) +
        geom_bar(fill="blue")
```

`summarise()` ungrouping output (override with `.groups` argument)



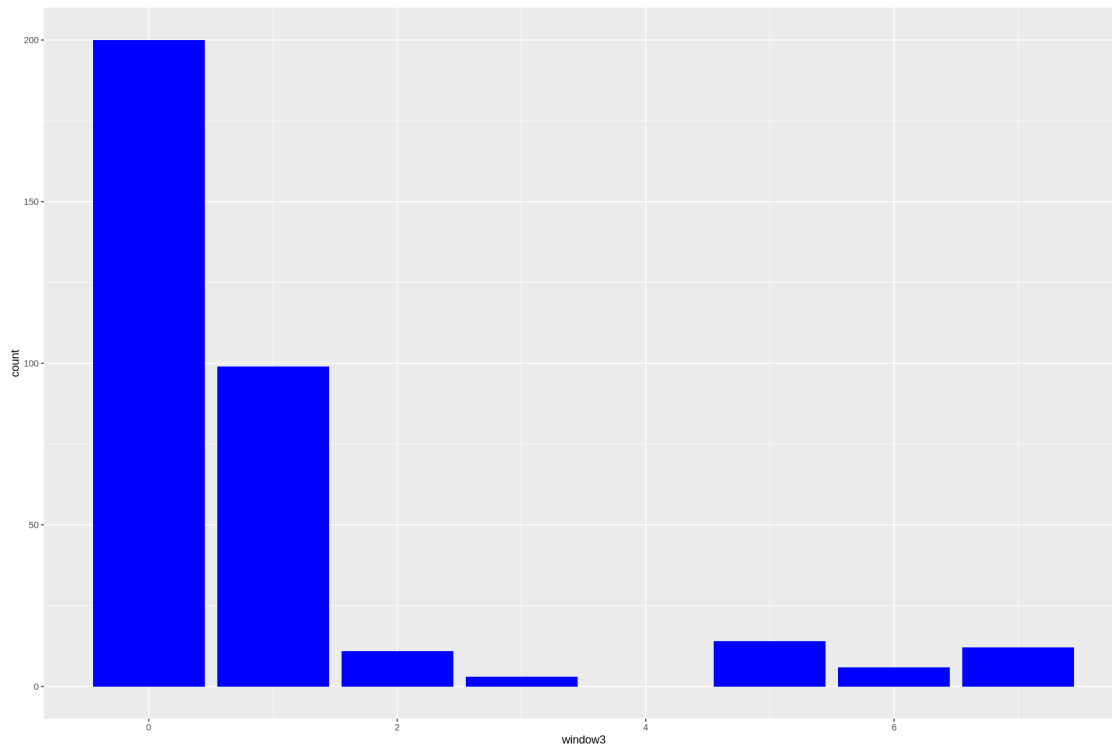
```
[44]: vlt %>%
      group_by(window2) %>%
      summarise(n = n()) %>%
      ggplot(aes(weight=n)) +
        aes(window2) +
        geom_bar(fill="blue")
```

`summarise()` ungrouping output (override with `.groups` argument)



```
[45]: vlt %>%
      group_by(window3) %>%
      summarise(n = n()) %>%
      ggplot(aes(weight=n)) +
        aes(window3) +
        geom_bar(fill="blue")
```

`summarise()` ungrouping output (override with `.groups` argument)



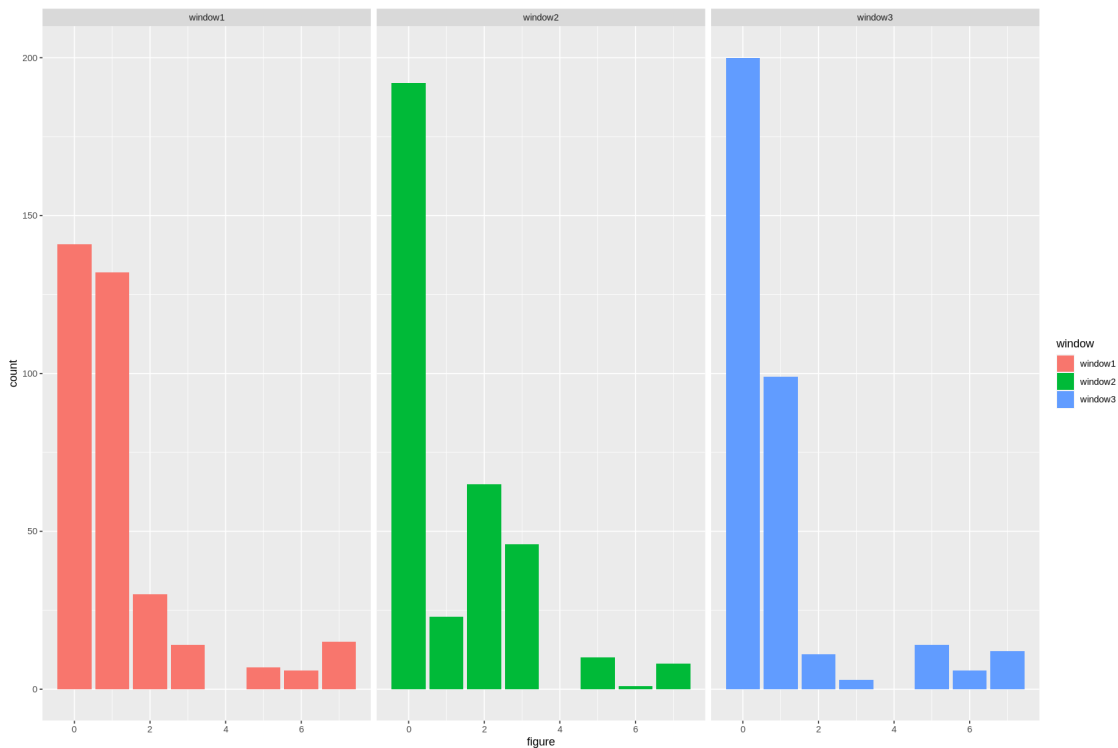
```
[46]: vlt %>%  
      pivot_longer(cols = starts_with('window'), names_to = "window", values_to = "figure")
```

prize <int>	night <int>	window <chr>	figure <int>
0	1	window1	2
0	1	window2	0
0	1	window3	0
0	1	window1	0
0	1	window2	5
0	1	window3	1
0	1	window1	0
0	1	window2	0
0	1	window3	0
0	1	window1	2
0	1	window2	0
0	1	window3	0
0	1	window1	0
0	1	window2	0
0	1	window3	0
0	1	window1	0
0	1	window2	0
0	1	window3	0
0	1	window1	0
0	1	window2	0
0	1	window3	1
0	1	window1	1
0	1	window2	0
0	1	window3	1
0	1	window1	0
0	1	window2	0
0	1	window3	0
5	1	window1	1
5	1	window2	2
5	1	window3	1
0	1	window1	0
0	1	window2	0
0	1	window3	1
0	2	window1	1
0	2	window2	0
0	2	window3	1
5	2	window1	1
5	2	window2	2
5	2	window3	3
0	2	window1	0
0	2	window2	2
0	2	window3	0
0	2	window1	0
0	2	window2	0
0	2	window3	0
0	2	window1	0
0	2	window2	1
0	2	window3	0
0	2	window1	7
0	2	window2	0
0	2	window3	0
0	2	window1	0
0	2	window2	0

A tibble: 1035 × 4


```
[47]: vlt %>%
      pivot_longer(cols = starts_with('window'), names_to = "window", values_to = "figure") %>%
      group_by(window, figure) %>%
      summarise(n = n()) %>%
      ggplot(aes(x = figure, weight = n, fill = window)) +
        facet_wrap(~window) +
        geom_bar()
```

`summarise()` regrouping output by 'window' (override with `.groups` argument)



Podemos observar que claramente no todos los simbolos tienen la misma frecuencia

4.3.2 Daño gastrointestinal

Del conjunto de datos **Lanza** del paquete **HSAUR2**, hay cuatr estudios, dibujen un gráfico para comparar los tamaños de los estudios, ¿son iguales? El resultado se mide en la variable clasificación, ¿qué opinan?

4.4 Ejercicios estructura

4.4.1 Películas:

¿Cómo se ve el scatterplot si quitamos los que tienen más de 1000 votos? Ahora filtremos los mayores a 9. ¿Pasa algo?

```
[58]: library(ggplot2movies)
      head(movies)
```

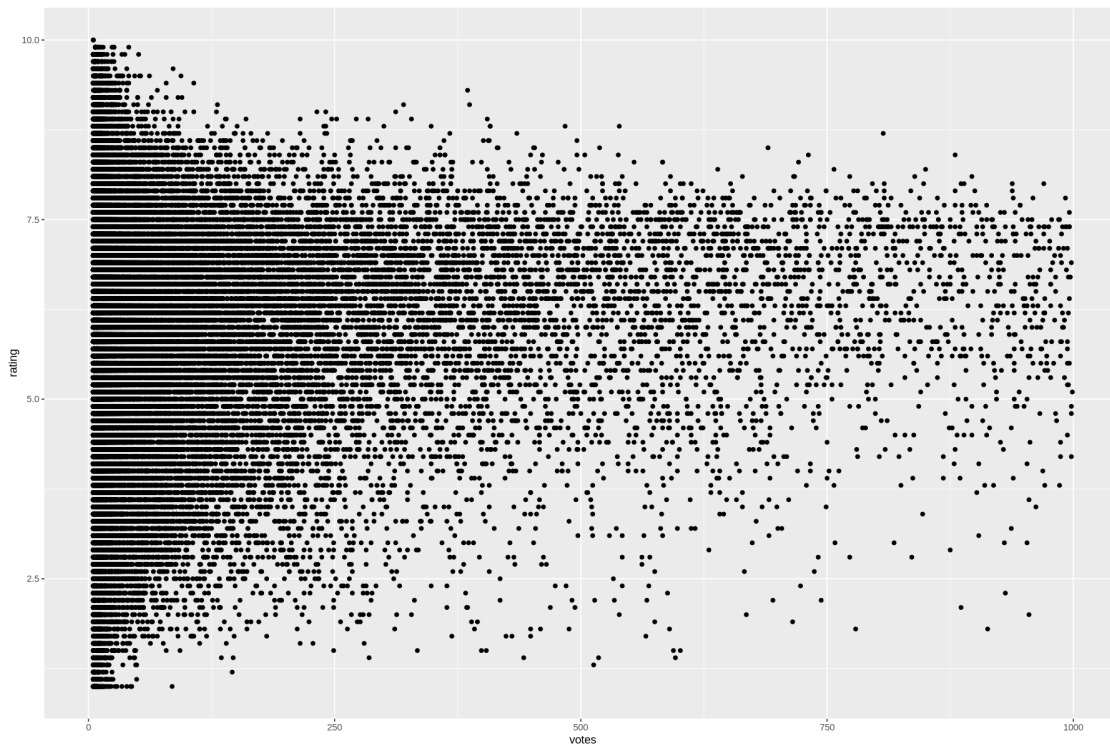
A tibble: 6 × 24

	title <chr>	year <int>	length <int>	budget <int>	rating <dbl>	votes <int>	r1 <dbl>	r2 <dbl>	r3 <dbl>
	\$	1971	121	NA	6.4	348	4.5	4.5	4.5
	\$1000 a Touchdown	1939	71	NA	6.0	20	0.0	14.5	4.5
	\$21 a Day Once a Month	1941	7	NA	8.2	5	0.0	0.0	0.0
	\$40,000	1996	70	NA	8.2	6	14.5	0.0	0.0
	\$50,000 Climax Show, The	1975	71	NA	3.4	17	24.5	4.5	0.0
	\$pent	2000	91	NA	4.3	45	4.5	4.5	4.5

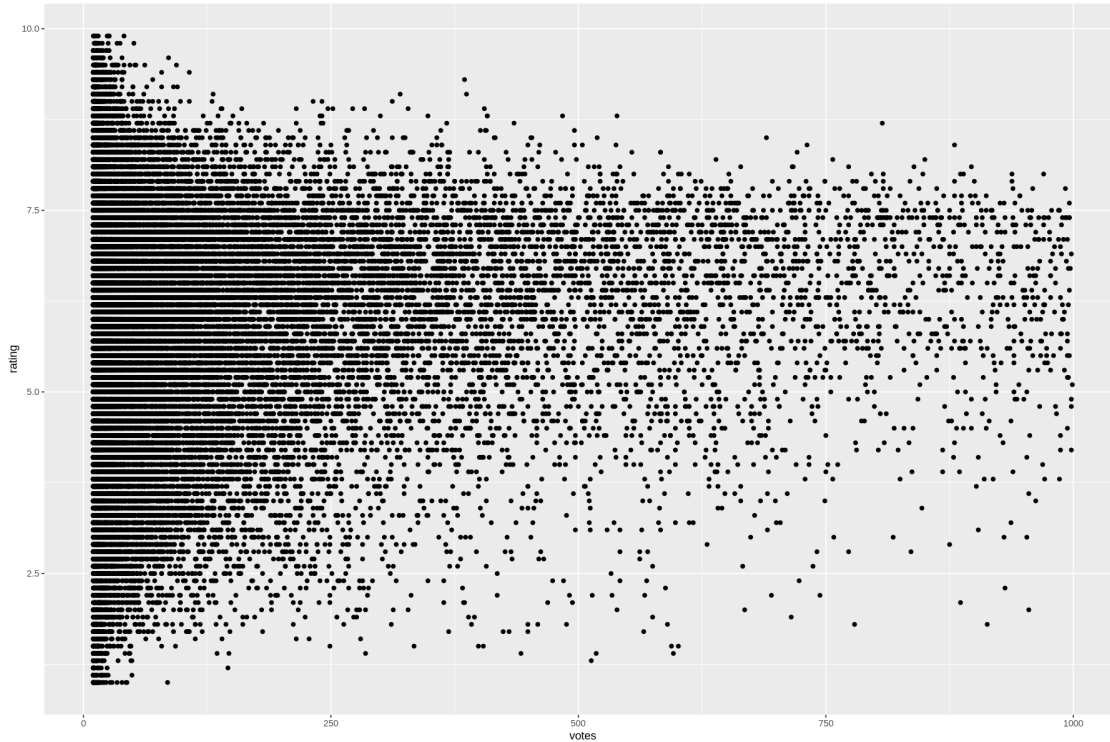
```
[59]: table(movies$mpaa)
```

	NC-17	PG	PG-13	R
53864	16	528	1003	3377

```
[60]: #Estamos quitando todo lo ruidoso (los de mas de 100 votos)
movies %>%
  filter(votes < 1000) %>%
  ggplot(aes(votes, rating)) +
    geom_point()
```



```
[61]: #Estamos quitando todo lo ruidoso (los de mas de 1000 votos y los de menos de 9)
movies %>%
  filter(votes < 1000, votes > 9) %>%
  ggplot(aes(votes, rating)) +
    geom_point()
```



4.4.2 Autos:

Grafiquen 1/MPG.City, vs horsepower ¿una relación lineal? ¿Cuáles son los outliers?

```
[48]: data(Cars93, package = "MASS")
      head(Cars93)
```

A data.frame: 6 × 27

	Manufacturer	Model	Type	Min.Price	Price	Max.Price	MPG.city	M
	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<int>	<i>
1	Acura	Integra	Small	12.9	15.9	18.8	25	31
2	Acura	Legend	Midsize	29.2	33.9	38.7	18	25
3	Audi	90	Compact	25.9	29.1	32.3	20	26
4	Audi	100	Midsize	30.8	37.7	44.6	19	26
5	BMW	535i	Midsize	23.7	30.0	36.2	22	30
6	Buick	Century	Midsize	14.2	15.7	17.3	22	31

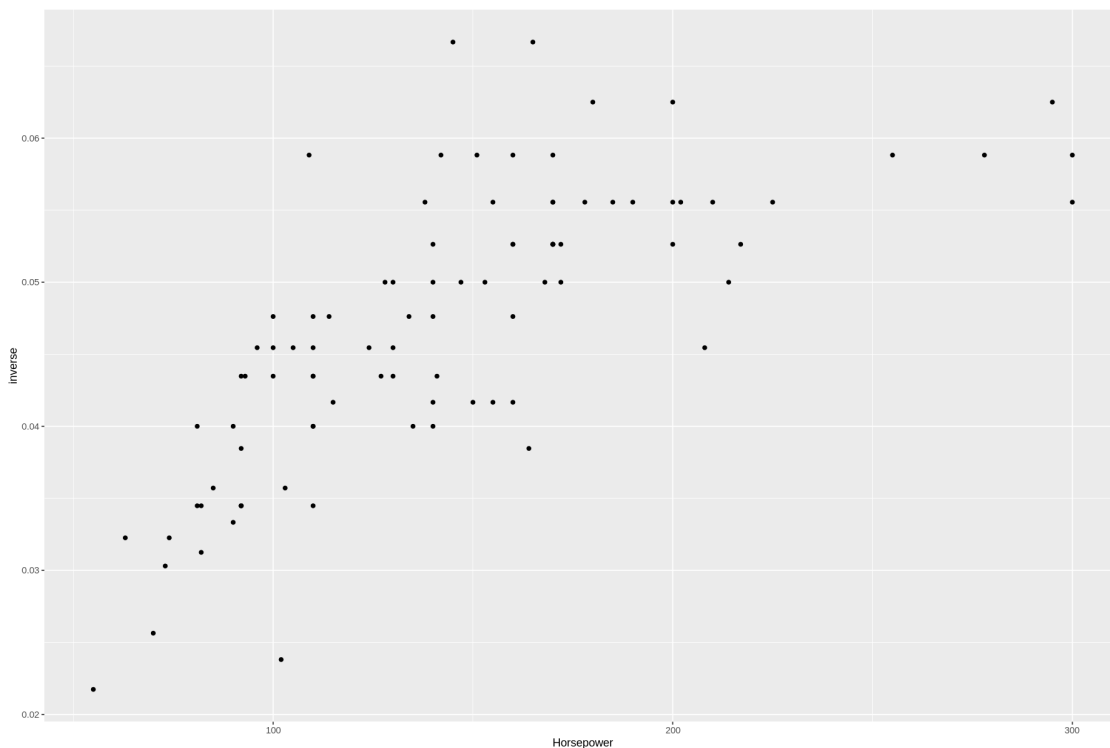
```
[50]: Cars93 %>%
      mutate(inverse = 1/MPG.city) %>%
      select(Horsepower, inverse) %>%
      ggplot(aes(x=Horsepower, inverse)) +
        geom_point()
```

```
Error in select(., Horsepower, inverse): unused arguments (Horsepower,  
↪ inverse)
```

Traceback:

1. Cars93 %>% mutate(inverse = 1/MPG.city) %>% select(Horsepower,
inverse) %>% ggplot(aes(x = Horsepower, inverse))
2. withVisible(eval(quote(`_fseq`(`_lhs`))), env, env))
3. eval(quote(`_fseq`(`_lhs`))), env, env)
4. eval(quote(`_fseq`(`_lhs`))), env, env)
5. `_fseq`(`_lhs`)
6. freduce(value, `_function_list`)
7. function_list[[i]](value)

```
[55]: Cars93 %>%  
mutate(inverse = 1/MPG.city) %>%  
ggplot(aes(x=Horsepower, inverse)) +  
geom_point()
```



Existe una relación directa entre el gasto de gasolina y los caballos de fuerza que tiene el motor. Aquellos que se encuentran separados en la parte superior derecha son autos deportivos.

4.4.3 Bancos

Dentro del paquete `Sleuth2`, los datos: `case1202`, tiene tres variables medidas en meses, de: Edad, Seniority, Experience.

¿Qué hay en la matriz de scatterplots? ¿Por?

5 Euler

Problema 13 y 16 del [Proyecto Euler](#)

6 Fecha de entrega

- Sabado 14 de Noviembre