Graficas-ejemplos

October 30, 2020

1 Velocidad de la luz

[9]:
```
library(ggplot2)
```

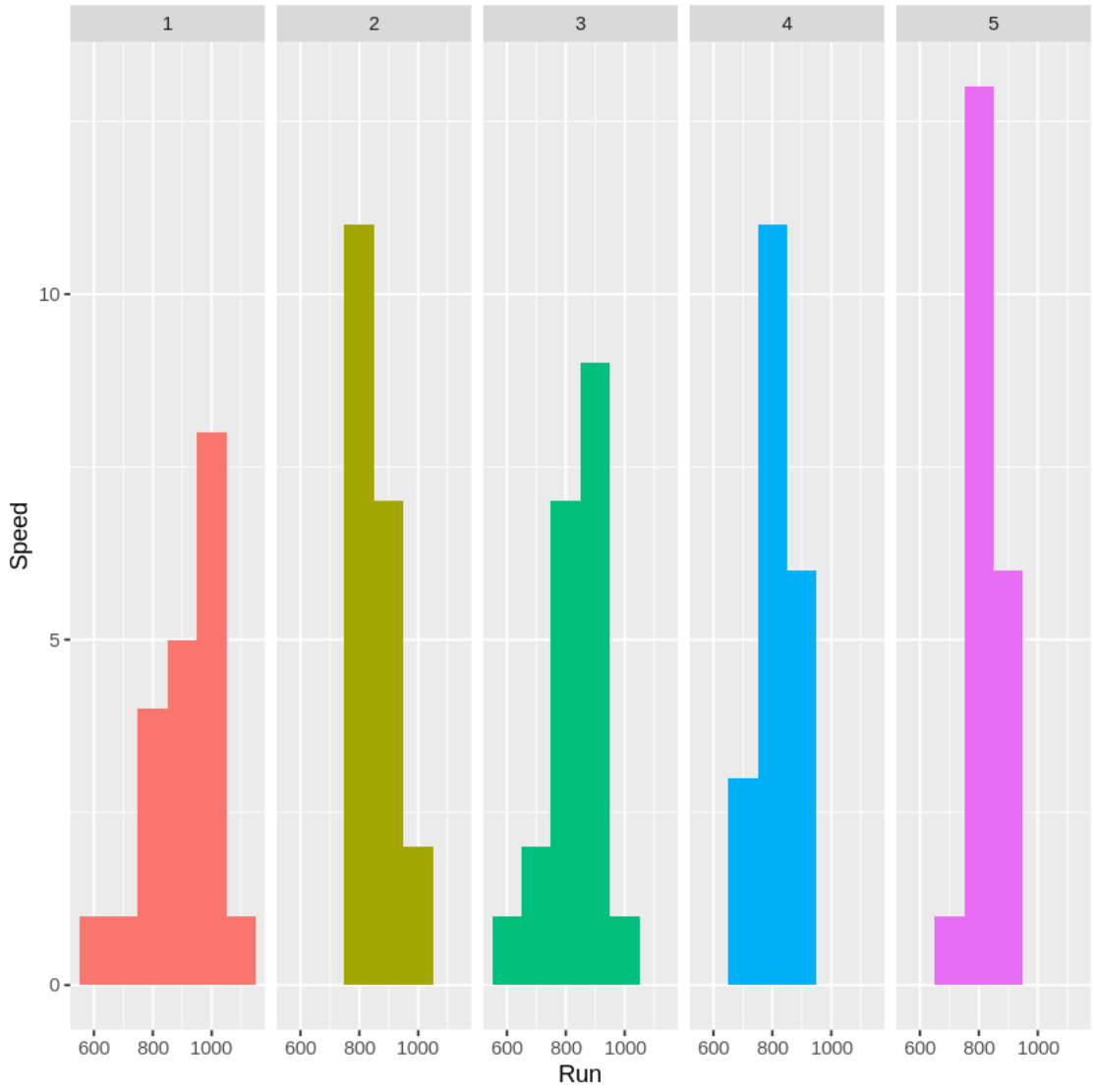
[10]:
```
library(MASS)
vl <- data.frame(michelson)
```

[11]:
```
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
A data.frame: 100 × 3 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

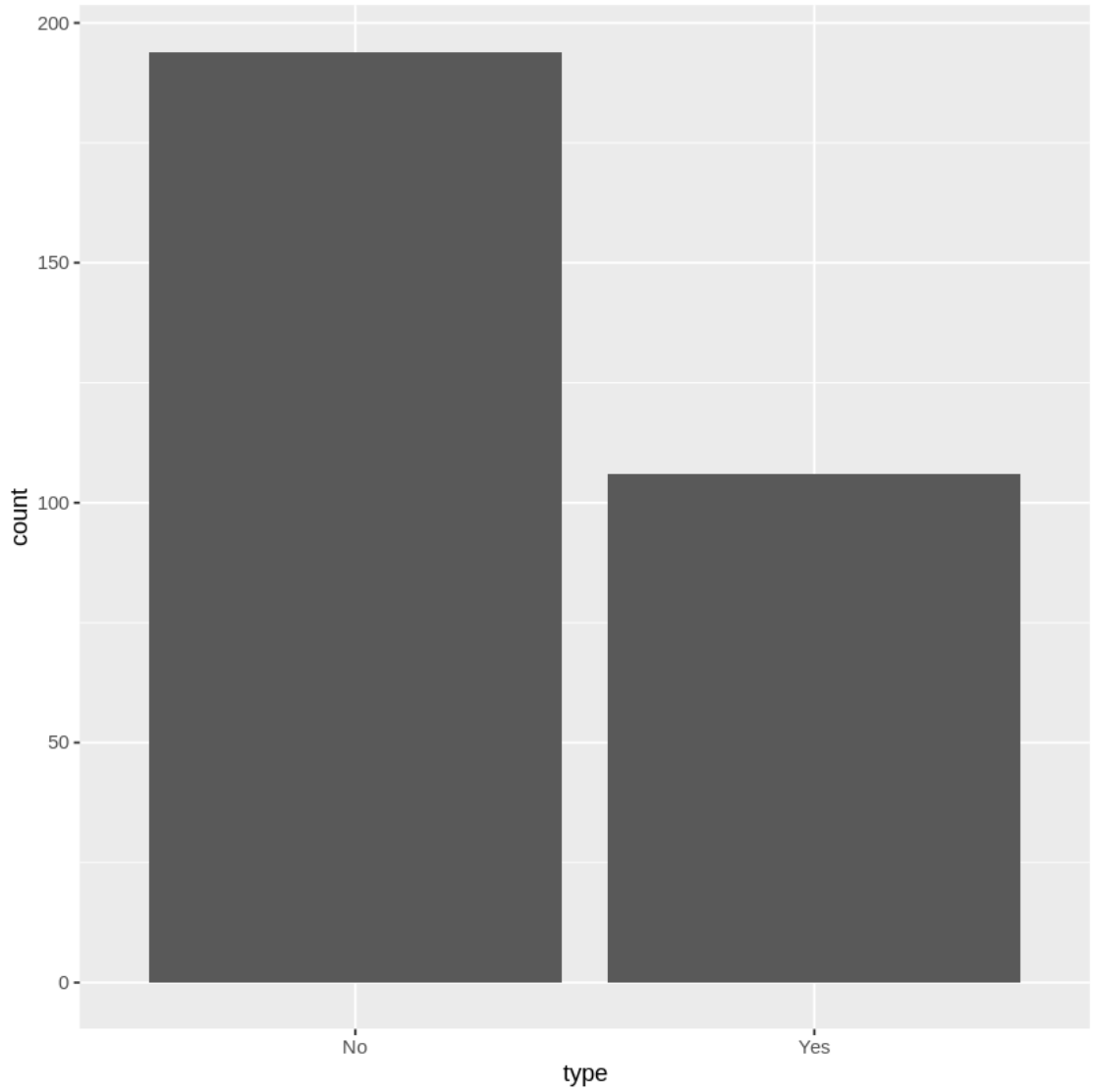
[15]:
```
ggplot(vl, aes(Speed, fill=Expt)) +
        geom_histogram(binwidth = 100) +
        xlab("Run") +
        ylab("Speed") +
        facet_grid(~Expt) +
        theme(legend.position = "none")


#aes --> graficamos speed con colores en funcion del experimento
#binwidth --> ancho de la barra
#facet_grid --> solo nos parta la grafica en funcion de Expt
    #Si quitamos fill se veria gris
```



2 Pima Indians

[17]:
```
ggplot(Pima.tr2, aes(type))+geom_bar()
```



[19]:
```
Pima.tr2
```

	npreg	glu	bp	skin	bmi	ped	age	type
	<int>	<int>	<int>	<int>	<dbl>	<dbl>	<int>	<fct>
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
A data.frame: 300×8 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

- ¿Qué dice este gráfico de barras?

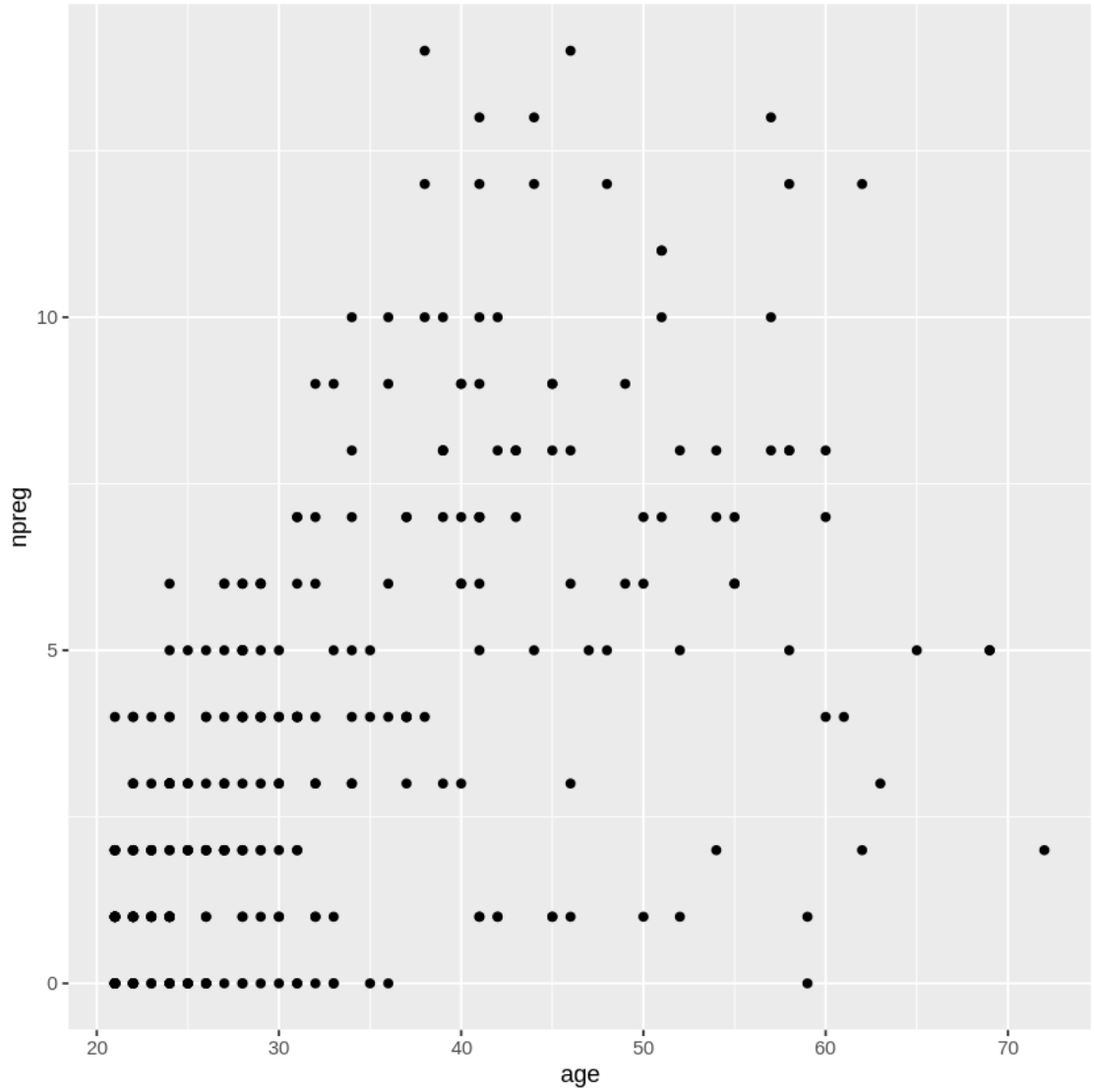
Las personas que tienen diabetes

[22]:
```
help(Pima.tr2)
```

- ¿Por qué la parte superior izquierda está vacía?

No hay muchas mujeres con muchos embarazos a los 20 años

[23]:
```
ggplot(Pima.tr2, aes(age,npreg))+geom_point()
```



3 Titanic

[25]:
```
head(data.frame(Titanic))
```

A data.frame: 6 × 5

	Class <fct>	Sex <fct>	Age <fct>	Survived <fct>	Freq <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

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

[29]:
```
library(tidyverse)
```

```
  Attaching packages                                    tidyverse
1.3.0

  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()
  dplyr::select() masks MASS::select()
```

[32]:
```
data.frame(Titanic) %>%
    group_by(Class) %>%
    summarise(count = sum(Freq))
```

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

A tibble: 4 × 2

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

4 Peliculas

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

[14]:
```
library(ggplot2)
install.packages("ggplot2movies")
library(ggplot2movies)
library(tidyverse)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done

  Attaching packages                                   tidyverse
1.3.0

 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()
```

[15]:
```
head(movies)
```

A tibble: 6 × 24

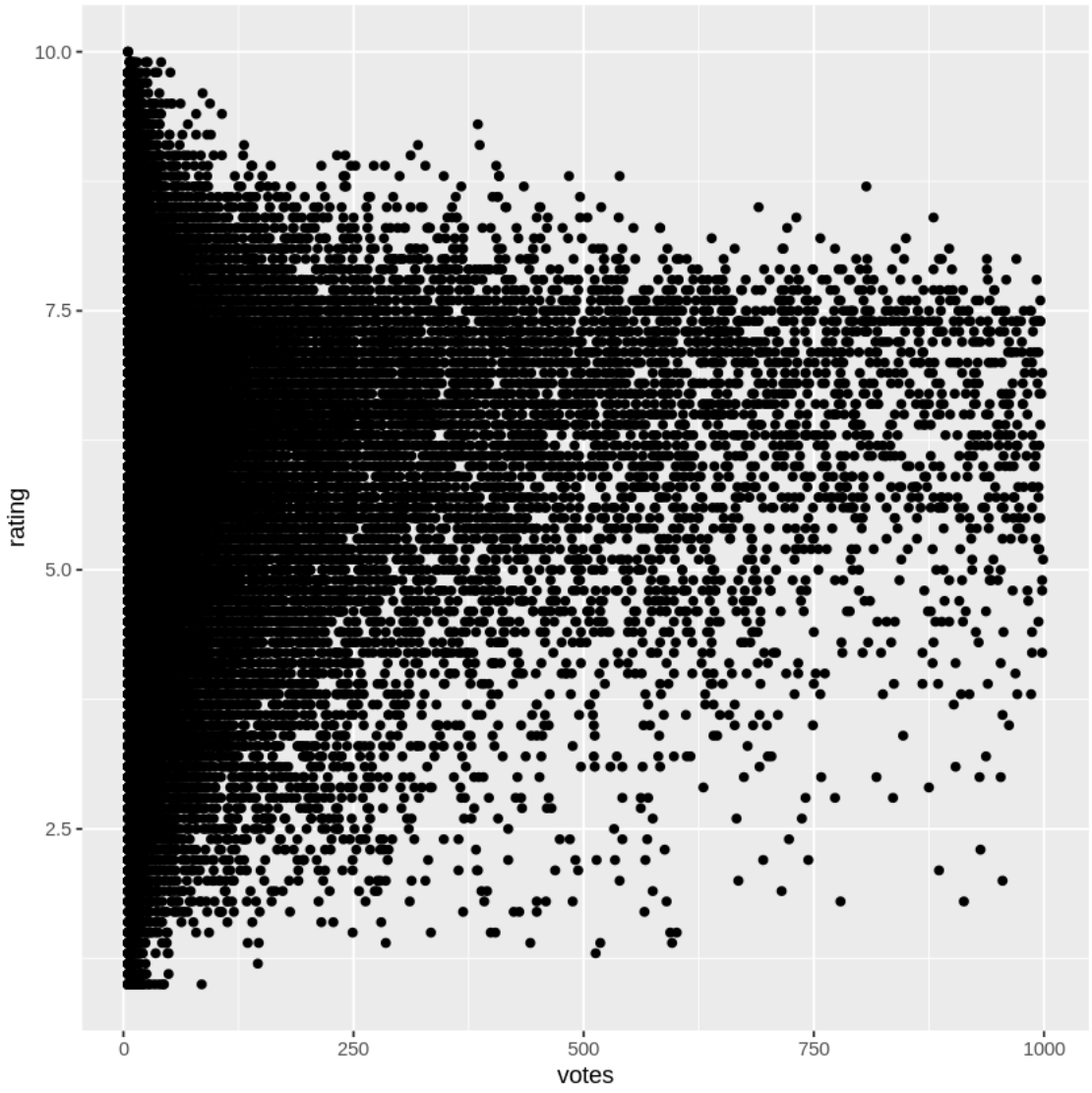
title <chr>	year <int>	length <int>	budget <int>	rating <dbl>	votes <int>	r1 <dbl>	r2 <dbl>	r3 <d
$	1971	121	NA	6.4	348	4.5	4.5	4.5
$1000 a Touchdown	1939	71	NA	6.0	20	0.0	14.5	4.5
$21 a Day Once a Month	1941	7	NA	8.2	5	0.0	0.0	0.0
$40,000	1996	70	NA	8.2	6	14.5	0.0	0.0
$50,000 Climax Show, The	1975	71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

[16]:
```
table(movies$mpaa)
```

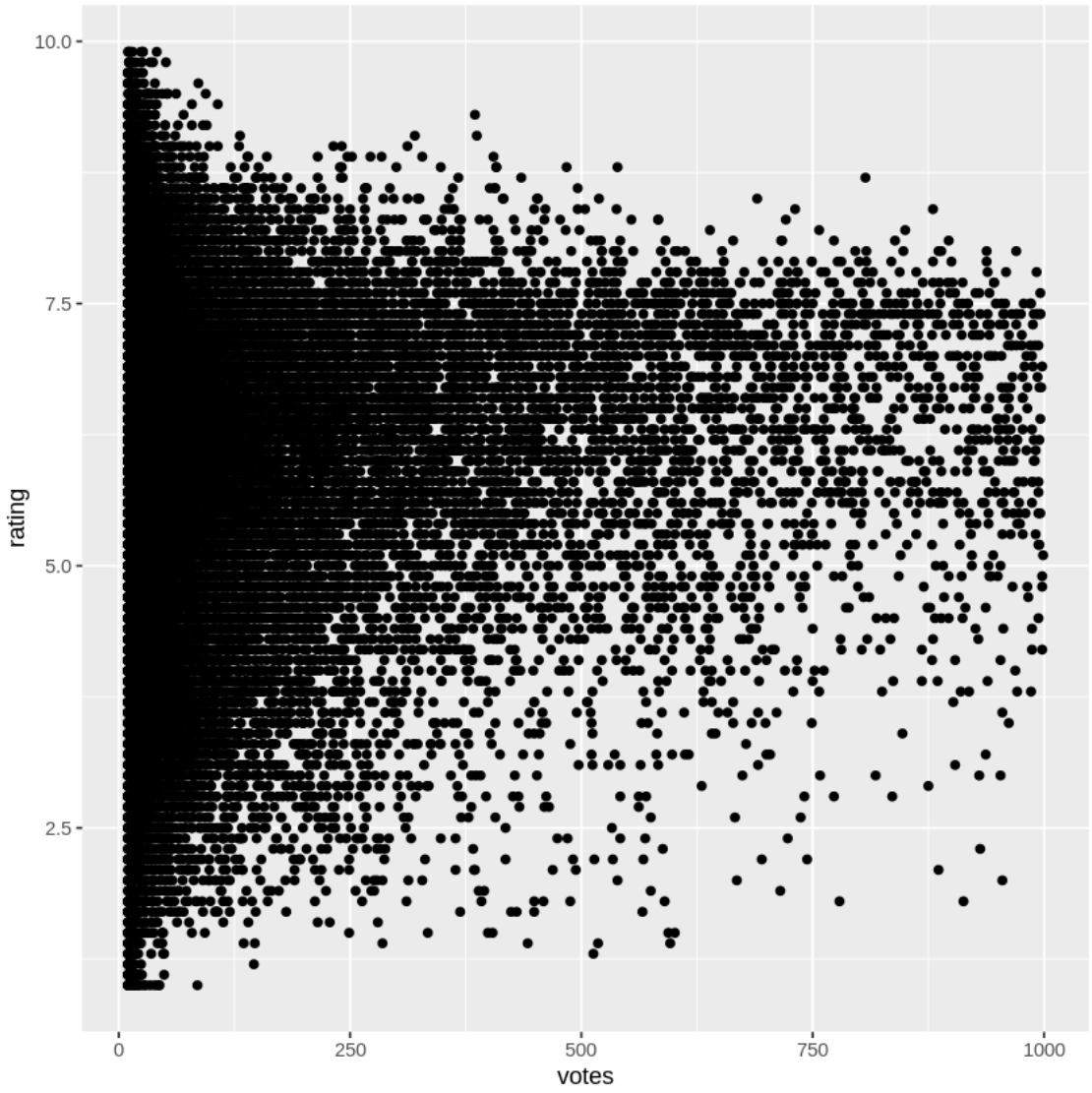
```
      NC-17    PG PG-13     R
53864    16   528  1003  3377
```

[17]:
```
#Estamos quitando todo lo ruidoso (los de mas de 100 votos)
movies %>%
    filter(votes < 1000) %>%
```

```
    ggplot(aes(votes, rating)) +
        geom_point()
```



[20]:
```
#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()
```



[39]:
```
movies %>%
    pivot_longer(cols = c(Action, Animation, Comedy, Drama, Documentary,␣
↪Romance, Short), names_to = "category", values_to = "figure") %>%
    data.frame() %>%
    head()
```

```
    Error in eval(lhs, parent, parent): object 'movies' not found
Traceback:


1. movies %>% pivot_longer(cols = c(Action, Animation, Comedy, Drama,
```

```
    .     Documentary, Romance, Short), names_to = "category", values_to =␣
↪"figure") %>%
    .     data.frame() %>% head()

       2. eval(lhs, parent, parent)

       3. eval(lhs, parent, parent)
```

https://tidyr.tidyverse.org/

5 Máquinas tragamonedas

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

[6]:
```
install.packages("DAAG")
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done
```

[10]:
```
library(tidyverse)
library(ggplot2)
```

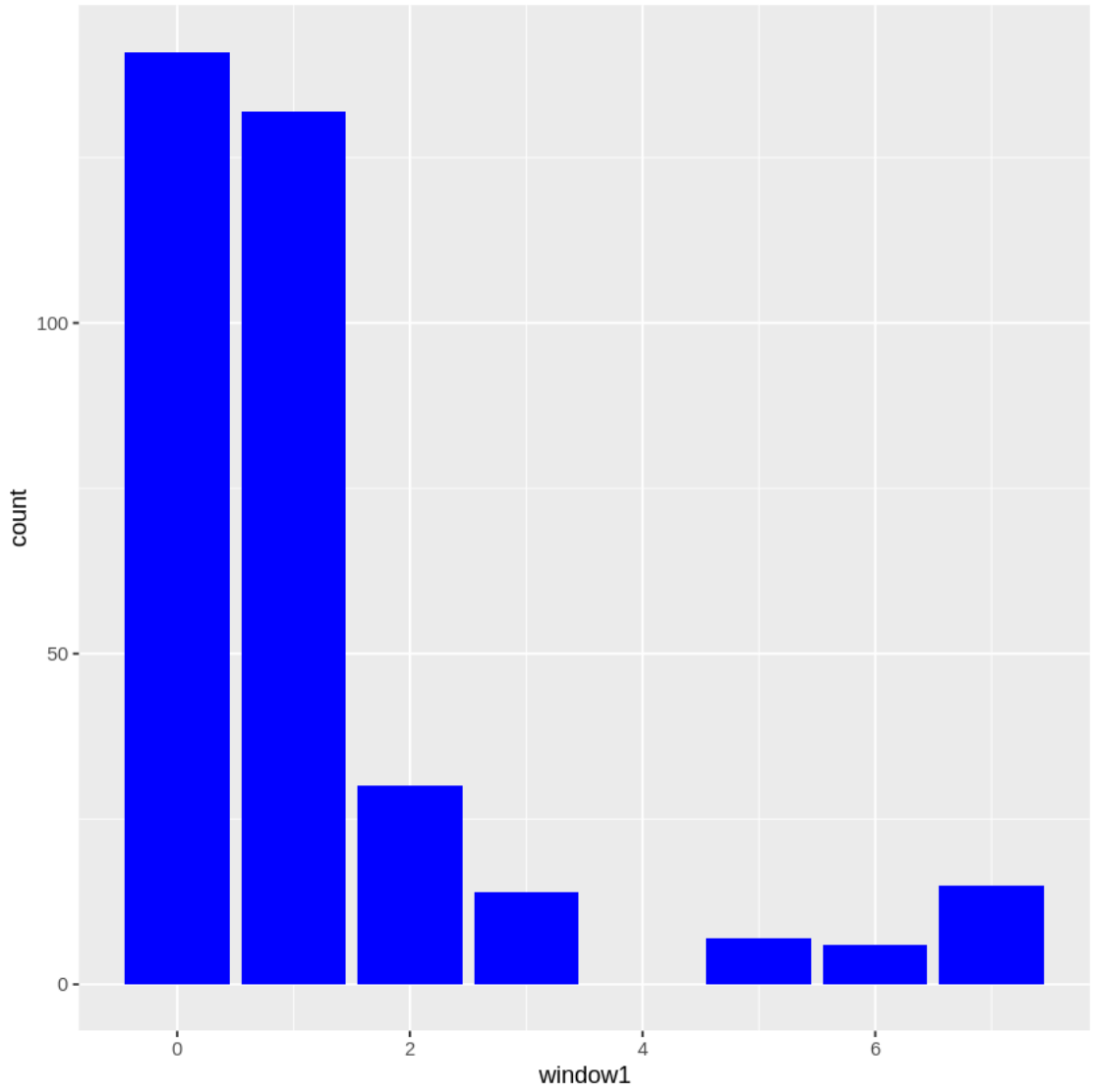
[8]:
```
data(vlt, package="DAAG")
head(vlt)
```

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

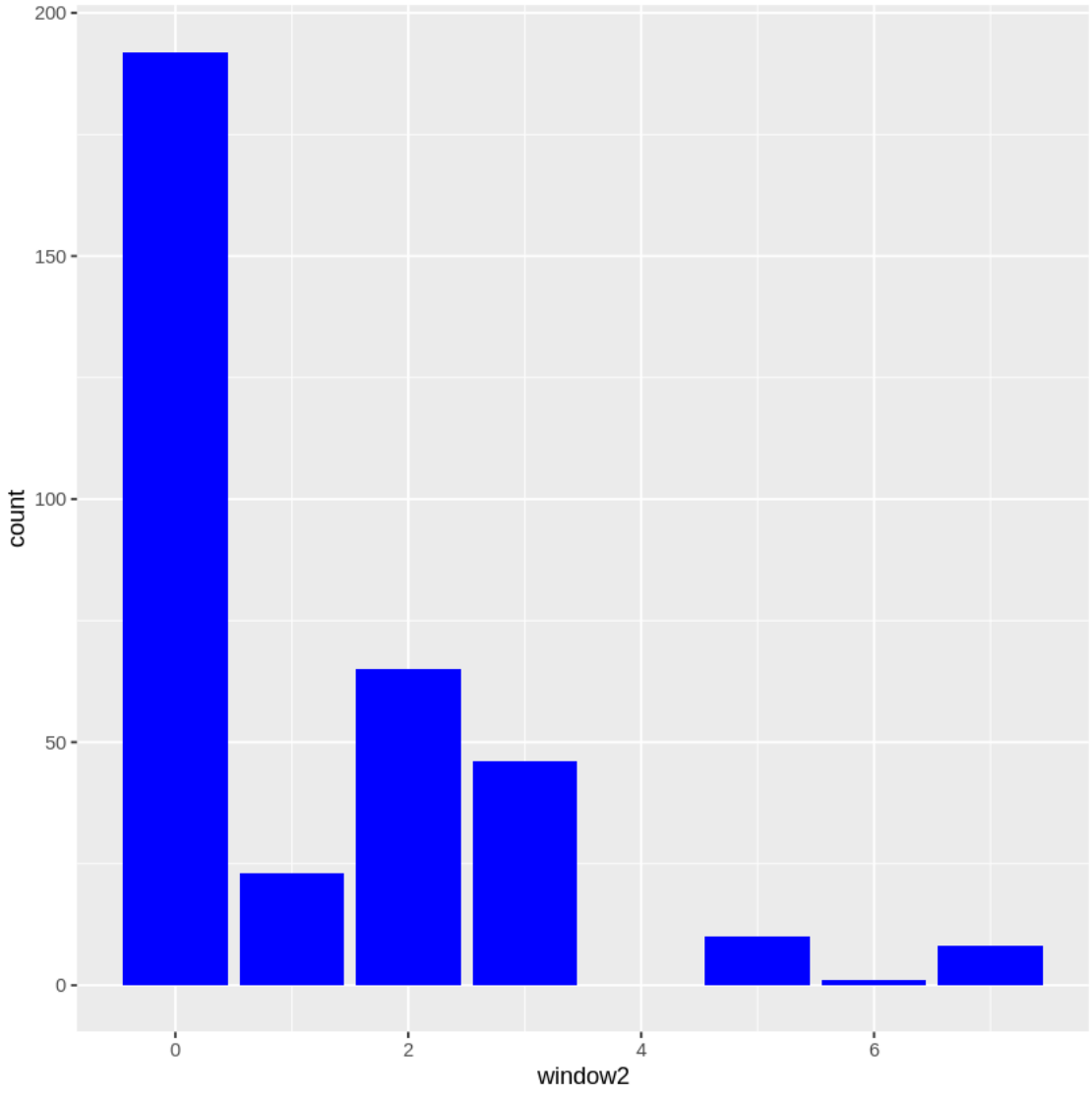
[21]:
```
vlt %>%
    group_by(window1) %>%
    summarise(n = n()) %>%
    ggplot(aes(weight=n)) +
        aes(window1) +
        geom_bar(fill="blue")
```

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



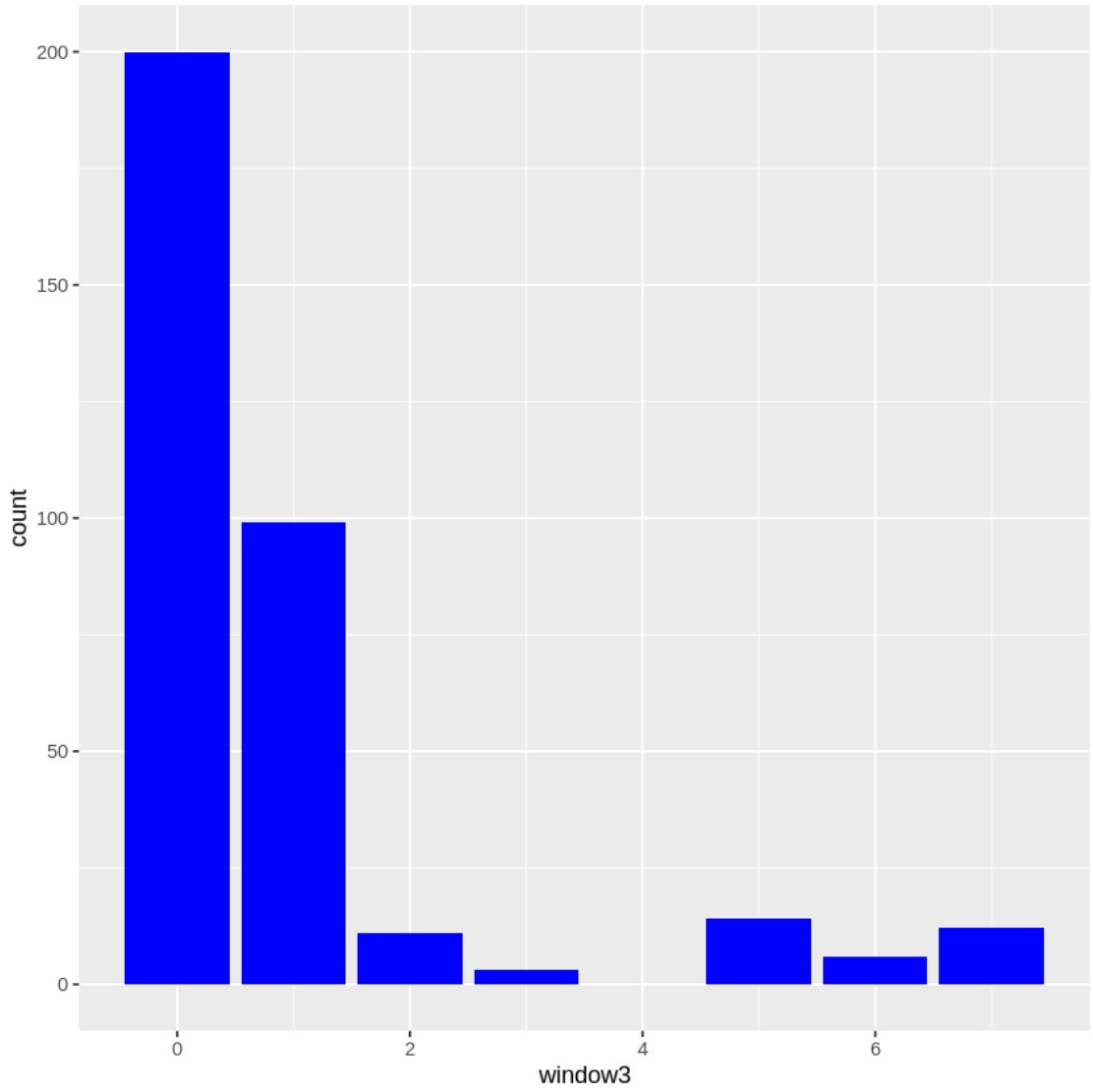
[19]:
```
vlt %>%
    group_by(window2) %>%
    summarise(n = n()) %>%
    ggplot(aes(weight=n)) +
        aes(window2) +
        geom_bar(fill="blue")
```

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



[18]:
```
vlt %>%
    group_by(window3) %>%
    summarise(n = n()) %>%
    ggplot(aes(weight=n)) +
        aes(window3) +
        geom_bar(fill="blue")
```

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



[23]:
```
#pivot_longer cambia el formato a una bonita
#names_to es el nombre de la columna
vlt %>%
    pivot_longer(cols = starts_with('window'), names_to = "window", values_to =␣
 ↪"figure" )
```

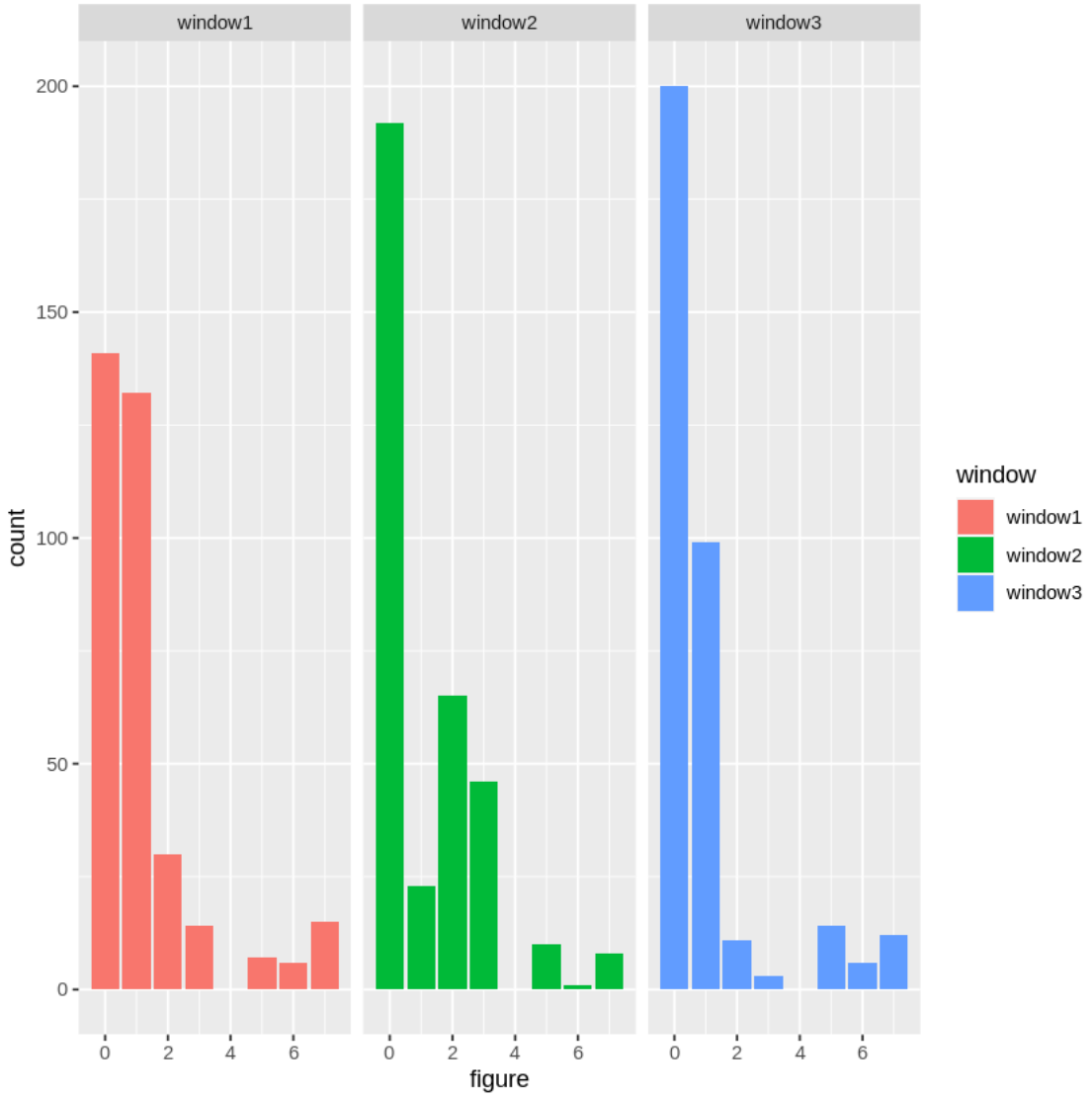
prize <int>	night <int>	window <chr>	figure <int>
0	1	window1	2
0	1	window2	0
0	1	window3	0
0	1	window1	0
0	1	window2	5
0	1	window3	1
0	1	window1	0
0	1	window2	0
0	1	window3	0
0	1	window1	2
0	1	window2	0
0	1	window3	0
0	1	window1	0
0	1	window2	0
0	1	window3	0
0	1	window1	0
0	1	window2	0
0	1	window3	1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

A tibble: 1035 × 4

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

[45]:
```
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)
```



6 Autos

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

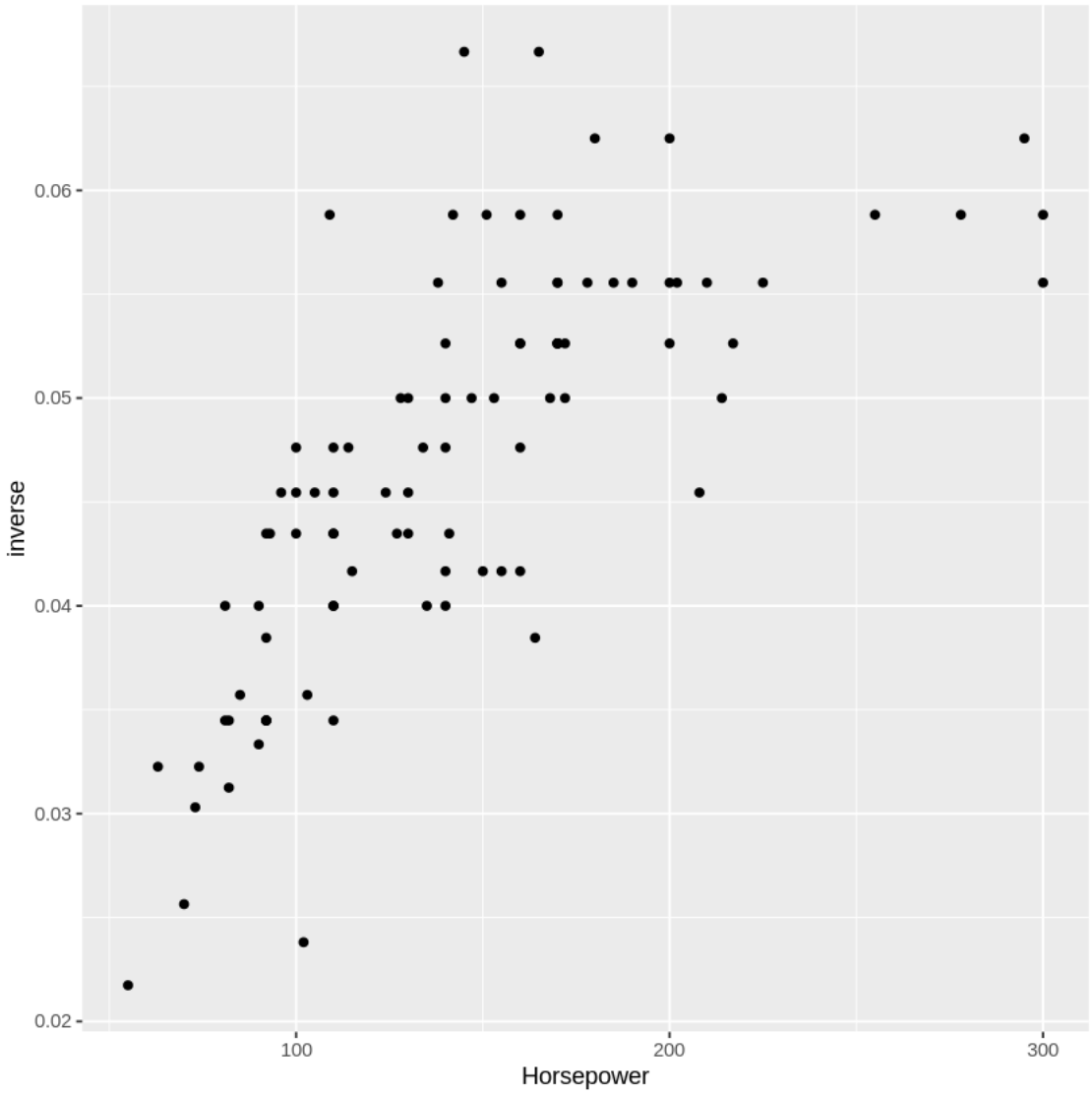
[41]:
```
data(Cars93, package = "MASS")
```

[43]:
```
head(Cars93)
```

A data.frame: 6 × 27

	Manufacturer <fct>	Model <fct>	Type <fct>	Min.Price <dbl>	Price <dbl>	Max.Price <dbl>	MPG.city <int>	MI <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

[47]:
```
Cars93 %>%
    mutate(inverse = 1/MPG.city) %>%
    select(Horsepower, inverse) %>%
    ggplot(aes(x=Horsepower, inverse)) +
        geom_point()
```


0.06
0.05
0.04
0.03
0.02
inverse
100
200
300
Horsepower