Data Visualization Analysis

Daven

2024-12-20

Table of contents

0.1 Introduction

This tutorial is designed to help you learn data visualization analysis by providing simple and useful information in a way that is easy to follow and understand.

0.2 Preparation

In order to draw a chart, we need to include the required packages for visualization and dataset. For example, `ggplot2` package is for drawing charts and `gcookbook` is for using `pg_mean` dataset.

```
library(ggplot2)
library(gcookbook)
```

0.3 Bar chart

In this section, we will draw a bar chart using `pg_mean` dataset. The dataset has two columns: group, `weight`.
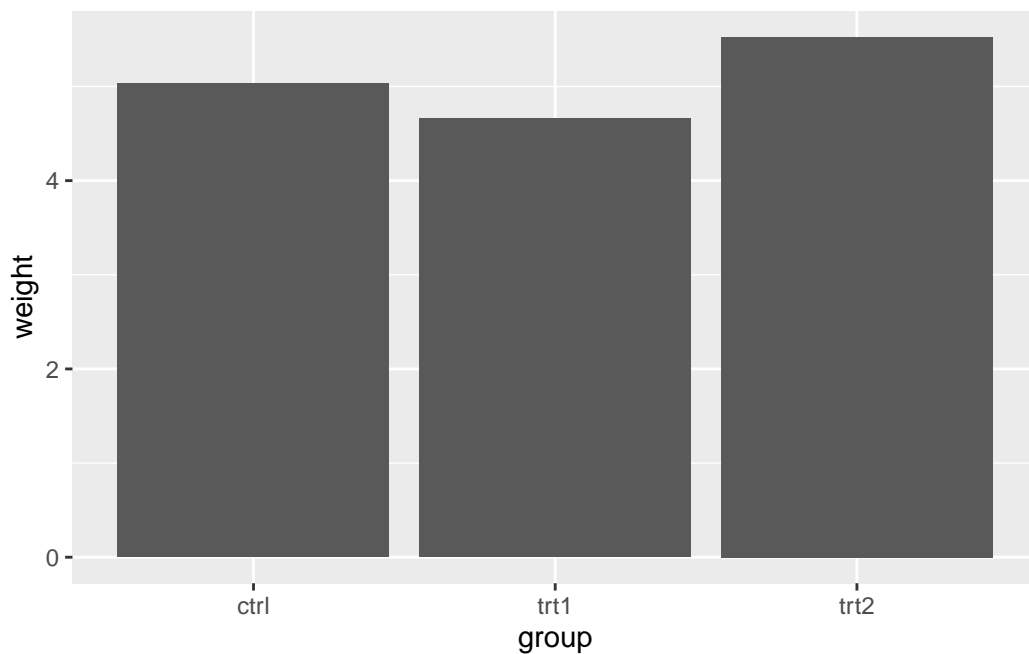
`pg_mean`

```
  group weight
1  ctrl  5.032
2  trt1  4.661
3  trt2  5.526
```

This dataset compares the weight across three groups:

- `ctrl`: Control group (baseline, weight = 5.032).
- `trt1`: Treatment 1 group (weight = 4.661).
- `trt2`: Treatment 2 group (weight = 5.526).

```
ggplot(pg_mean, aes(x = group, y = weight)) +
  geom_col()
```


It initializes a ggplot with the dataset `pg_mean`.

`aes(x = group, y = weight)` specifies the aesthetics:

- x = group: Assign the `group` variable to the x-axis (categorical data, such as `ctrl`, `trt1`, `trt2`).
- y = weight: Assign the `weight` variable to the y-axis (numerical data).

`geom_col():`

- Adds a column geometry to the plot.
- `geom_col()` creates bars where the height of each bar corresponds to the value of weight for each group.

0.4 Bar chart with color

We can expand further our bar chart by adding colors to its shape and outline to add more visualization into it. However, first we need to include **gcookbook** for using **pg_mean** dataset and **ggplot2** for drawing the chart.

```
library(gcookbook)
library(ggplot2)
```
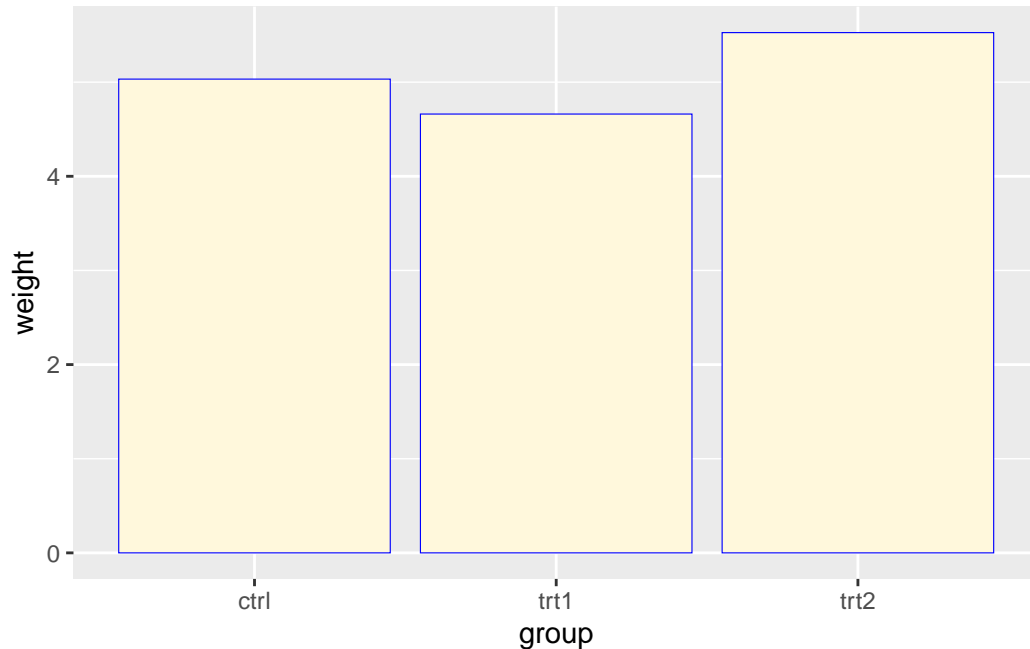
Here we need to initializes **pg_mean** dataset by using **gcookbook** library within our code.

```
pg_mean
```

```
  group weight
1  ctrl  5.032
2  trt1  4.661
3  trt2  5.526
```

A quick preview of **pg_mean** dataset.

```
ggplot(pg_mean, aes(x = group, y = weight)) +
  geom_col(fill = 'cornsilk', colour = 'blue', size = .2)
```

As we demonstrate before at Bar chart, `aes(x = group, y = weight)` specifies the aesthetics:

- `x = group`: Assign the `group` variable to the x-axis (categorical data, such as `ctrl`, `trt1`, `trt2`).
- `y = weight`: Assign the `weight` variable to the y-axis (numerical data).

`geom_col():`

- Adds a column geometry to the plot.
- `geom_col()` creates bars where the height of each bar corresponds to the value of weight for each group.

Inside `geom_col()` command, `fill = "cornsilk", colour = "blue", size = .2` specifies the following:

- `fill = "cornsilk"`: Assign color to the bar's area using the color `"cornsilk"`.
- `colour = "blue"`: Assign color to the bar's outline using the color `"blue"`.
- `size = .2`: Specified the size of the outline for all the bar within the chart.

```
labs(title = 'Bar Chart w/ Color',
      caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

4

To add more details on your chart, we can add `labs()` and `theme()` into our codes.

`labs()`:

- Used to customized axis labels and legend titles.
- To add `title` and `caption` unto the chart.

`theme()`:

- Used to give plots a consistent customized look.
- Positioning your chart `title` and adjusting the `size` of your chart font

```
labs(title = 'Bar Chart w/ Color',
     caption = 'By Daven, DV, THU, 2024')
```

Inside `labs()`:

- `title = 'Bar Chart w/ Color'`: To add "Bar Chart w/ Color" as the `title`.
- `caption = 'By Daven, DV, THU, 2024'`: To add "By Daven, DV, THU, 2024" as the `caption`.

```
theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Inside `theme()`:

- `plot.title = element_text()`: To customize the `title` chart to your desire output.

Inside `plot.title = element_text()`:

- `hjust = 0.5`: Positioning your chart `title` in the middle.
- `size = 20`: Sizing your title text to `size` 20.

```
library(gcookbook)
library(ggplot2)

ggplot(pg_mean, aes(x = group, y = weight)) +
  geom_col(fill = 'cornsilk', colour = 'blue', size = .2) +
  labs(title = 'Bar Chart',
       caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Bar Chart
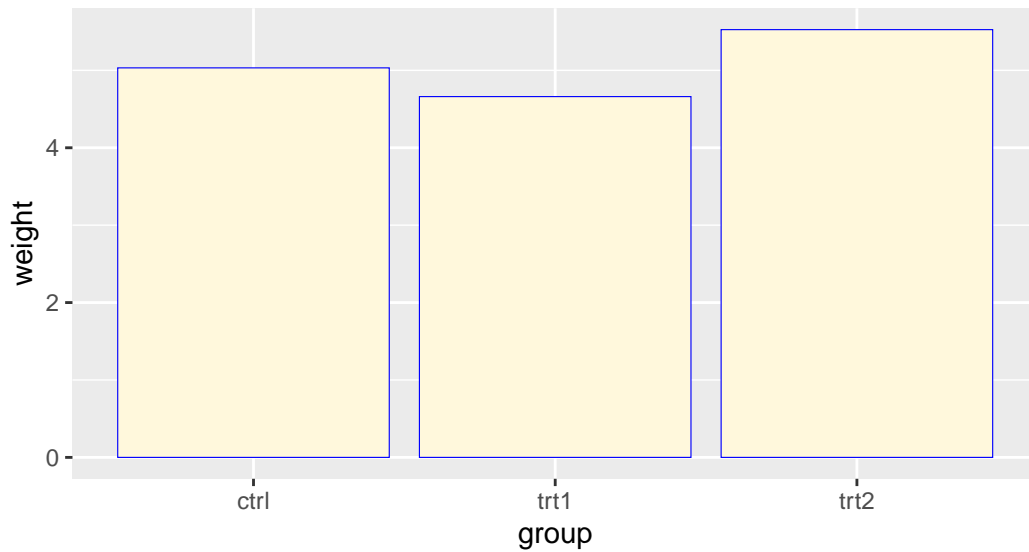
For a better view of the overall explanation above, this is the complete structure on how the codes *given above* should be arranged.

Note: You can use different colors on chart's area and outline like `'red'` or `'green'` as well as the size of your chart title and positioning (`hjust = 1, size = 15`).

0.5 Line chart

Using the same Preparation, we'll be using the same library which is `gcookbook` and `ggplot2` library.

```
library(gcookbook)
library(ggplot2)
```

Here we need to initializes `BOD` dataset by using `gcookbook()` library within our code.

```
BOD
```

```
  Time demand
1    1    8.3
2    2   10.3
3    3   19.0
```

```
4     4    16.0
5     5    15.6
6     7    19.8
```

A quick preview of `BOD` dataset.

```
ggplot(BOD, aes(x = Time, y = demand)) +
  geom_line()
```
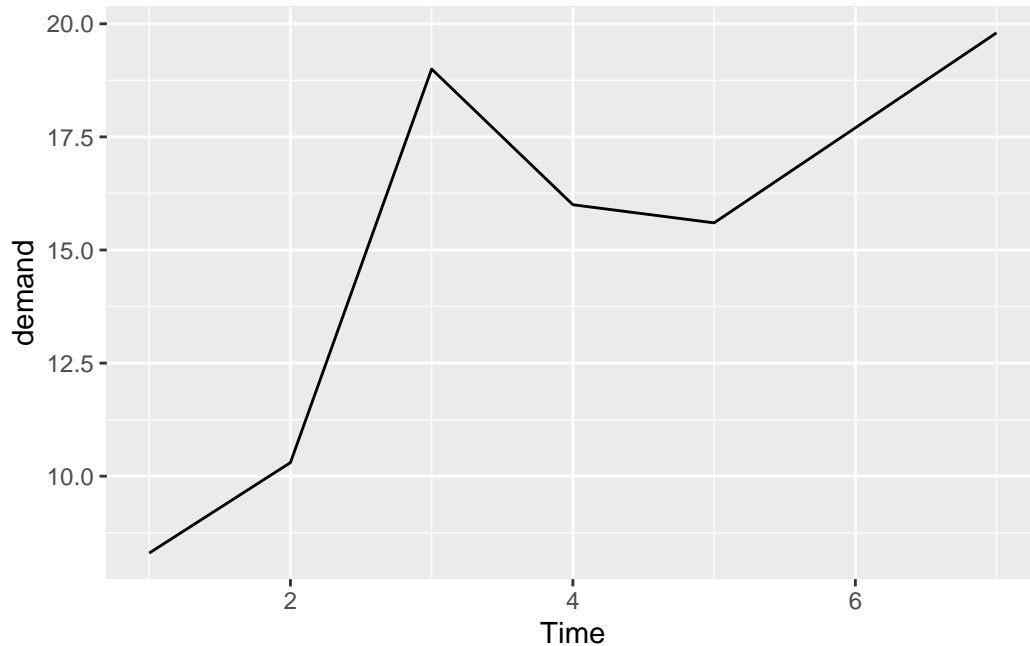

In this part of chart, `aes(x = Time, y = demand)` specifies the aesthetics:

- `x = Time`: Assign the `Time` variable to the x-axis.
- `y = demand`: Assign the 'demand variable to the y-axis (numerical data).

`geom_line`:

- Adds a line geometry to the plot.
- `geom_line` creates lines that connect data points over a continuous variable.

```
labs(title = 'Line Chart',
       caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

To add more details on your chart, we can add `labs()` and `theme()` into our codes.

`labs()`:

- Used to customized axis labels and legend titles.
- To add `title` and `caption` unto the chart.

`theme()`:

- Used to give plots a consistent customized look.
- Positioning your chart `title` and adjusting the `size` of your chart font.

```r
labs(title = 'Line Chart',
     caption = 'By Daven, DV, THU, 2024')
```

Inside `labs()`:

- `title = 'Line Chart'`: To add "Line Chart" as the `title`.
- `caption = 'By Daven, DV, THU, 2024'`: To add "By Daven, DV, THU, 2024" as the `caption`.

```r
theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Inside `theme()`:

- `plot.title = element_text()`: To customize the `title` chart to your desire output.

Inside `plot.title = element_text()`:

- `hjust = 0.5`: Positioning your chart `title` in the middle.
- `size = 20`: Sizing your title text to `size` 20.

```r
library(gcookbook)
library(ggplot2)

ggplot(BOD, aes(x = Time, y = demand)) +
  geom_line() +
  labs(title = 'Line Chart',
       caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Line Chart
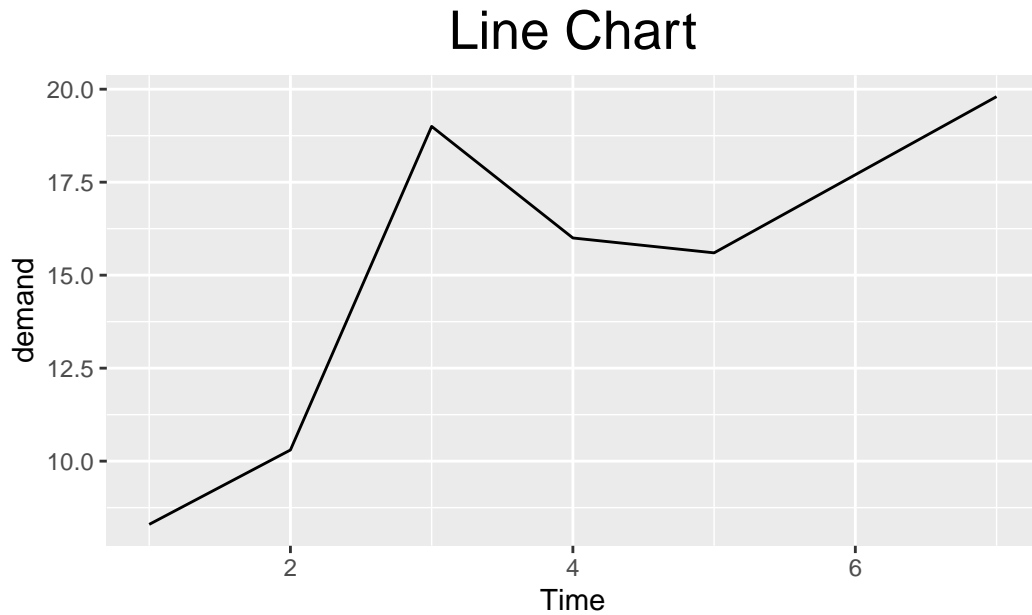
By Daven, DV, THU, 2024

For a better view of the overall explanation above, this is the complete structure on how the codes *given above* should be arranged.

Note: You can use different size of your chart title and positioning (`hjust = 1, size = 15`).

0.6 Histogram

In this part we'll be drawing the `diamonds` dataset as a histogram chart.

```
library(ggplot2)
```

To start, we'll be only using the `gcookbook` library to use the `diamonds` dataset.

```
diamonds
```

```
# A tibble: 53,940 x 10
   carat cut       color clarity depth table price     x     y     z
   <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
 1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
 2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
 3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31
```

9

```
 4  0.29 Premium   I    VS2     62.4   58   334  4.2   4.23  2.63
 5  0.31 Good      J    SI2     63.3   58   335  4.34  4.35  2.75
 6  0.24 Very Good J    VVS2    62.8   57   336  3.94  3.96  2.48
 7  0.24 Very Good I    VVS1    62.3   57   336  3.95  3.98  2.47
 8  0.26 Very Good H    SI1     61.9   55   337  4.07  4.11  2.53
 9  0.22 Fair      E    VS2     65.1   61   337  3.87  3.78  2.49
10  0.23 Very Good H    VS1     59.4   61   338  4     4.05  2.39
# i 53,930 more rows
```

A quick preview of the `diamonds` dataset.

```
ggplot(diamonds, aes (x = carat)) +
  geom_histogram()
```
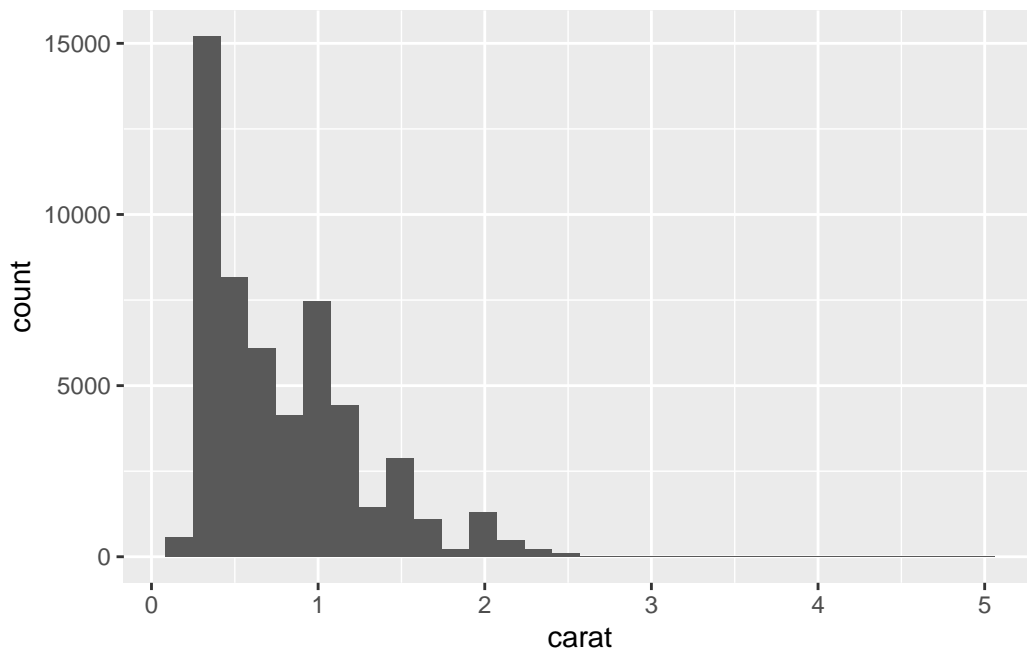

in this part of chart, `aes (x = carat)` specifies the aesthetics:

- `x = carat`: Assign the `carat` variable to the x-axis.

`geom_histogram()`:

- Adds a frequency of column geometry similar to a bar to the plot.
- `geom_histogram()` visualizes the distribution of a continuous variable by dividing it into bins.

```
labs(title = 'Histogram Chart',
        caption = 'By Daven, DV, THU') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

To add more details on your chart, we can add `labs()` and `theme()` into our codes.

`labs()`:

- Used to customized axis labels and legend titles.
- To add `title` and `caption` unto the chart.

`theme()`:

- Used to give plots a consistent customized look.
- Positioning your chart `title` and adjusting the `size` of your chart font.

```
labs(title = 'Histogram Chart',
        caption = 'By Daven, DV, THU')
```

Inside `labs()`:

- `title = 'Histogram Chart'`: To add "Histogram Chart" as the `title`.
- `caption = 'By Daven, DV, THU, 2024'`: To add "By Daven, DV, THU, 2024" as the `caption`.

```
theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Inside `theme()`:

- `plot.title = element_text()`: To customize the `title` chart to your desire output.

Inside `plot.title = element_text()`:

- `hjust = 0.5`: Positioning your chart `title` in the middle.
- `size = 20`: Sizing your title text to `size` 20.

```
library(ggplot2)

ggplot(diamonds, aes (x = carat)) +
  geom_histogram() +
  labs(title = 'Histogram Chart',
        caption = 'By Daven, DV, THU') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Histogram Chart
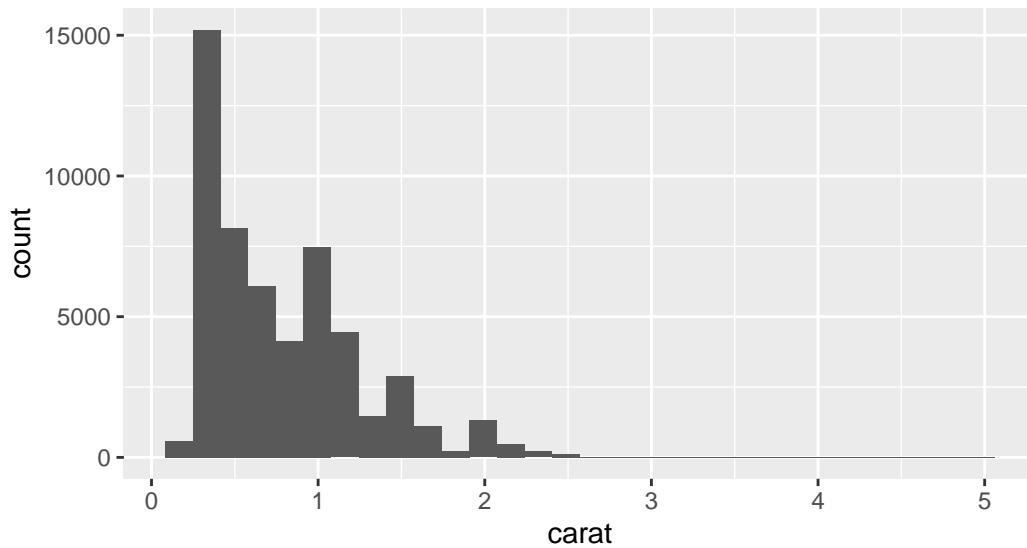
By Daven, DV, THU

For a better view of the overall explanation above, this is the complete structure on how the codes *given above* should be arranged.

Note: You can use different size of your chart title and positioning (`hjust = 1, size = 15`).

0.7 Correlation chart

In this part, we're going to make correlation/scatter chart using `heightweight` dataset.

```
library(gcookbook)
library(ggplot2)
```

Here we need to initialize the `heighweight` dataset using `gcookbook` library.

```
heightweight
```

```
  sex ageYear ageMonth heightIn weightLb
1   f   11.92      143     56.3     85.0
2   f   12.92      155     62.3    105.0
3   f   12.75      153     63.3    108.0
4   f   13.42      161     59.0     92.0
```

5	f	15.92	191	62.5	112.5
6	f	14.25	171	62.5	112.0
7	f	15.42	185	59.0	104.0
8	f	11.83	142	56.5	69.0
9	f	13.33	160	62.0	94.5
10	f	11.67	140	53.8	68.5
11	f	11.58	139	61.5	104.0
12	f	14.83	178	61.5	103.5
13	f	13.08	157	64.5	123.5
14	f	12.42	149	58.3	93.0
15	f	11.92	143	51.3	50.5
16	f	12.08	145	58.8	89.0
17	f	15.92	191	65.3	107.0
18	f	12.50	150	59.5	78.5
19	f	12.25	147	61.3	115.0
20	f	15.00	180	63.3	114.0
21	f	11.75	141	61.8	85.0
22	f	11.67	140	53.5	81.0
23	f	13.67	164	58.0	83.5
24	f	14.67	176	61.3	112.0
25	f	15.42	185	63.3	101.0
26	f	13.83	166	61.5	103.5
27	f	14.58	175	60.8	93.5
28	f	15.00	180	59.0	112.0
29	f	17.50	210	65.5	140.0
30	f	12.17	146	56.3	83.5
31	f	14.17	170	64.3	90.0
32	f	13.50	162	58.0	84.0
33	f	12.42	149	64.3	110.5
34	f	11.58	139	57.5	96.0
35	f	15.50	186	57.8	95.0
36	f	16.42	197	61.5	121.0
37	f	14.08	169	62.3	99.5
38	f	14.75	177	61.8	142.5
39	f	15.42	185	65.3	118.0
40	f	15.17	182	58.3	104.5
41	f	14.42	173	62.8	102.5
42	f	13.83	166	59.3	89.5
43	f	14.00	168	61.5	95.0
44	f	14.08	169	62.0	98.5
45	f	12.50	150	61.3	94.0
46	f	15.33	184	62.3	108.0
47	f	11.58	139	52.8	63.5

48	f	12.25	147	59.8	84.5
49	f	12.00	144	59.5	93.5
50	f	14.75	177	61.3	112.0
51	f	14.83	178	63.5	148.5
52	f	16.42	197	64.8	112.0
53	f	12.17	146	60.0	109.0
54	f	12.08	145	59.0	91.5
55	f	12.25	147	55.8	75.0
56	f	12.08	145	57.8	84.0
57	f	12.92	155	61.3	107.0
58	f	13.92	167	62.3	92.5
59	f	15.25	183	64.3	109.5
60	f	11.92	143	55.5	84.0
61	f	15.25	183	64.5	102.5
62	f	15.42	185	60.0	106.0
63	f	12.33	148	56.3	77.0
64	f	12.25	147	58.3	111.5
65	f	12.83	154	60.0	114.0
66	f	13.00	156	54.5	75.0
67	f	12.00	144	55.8	73.5
68	f	12.83	154	62.8	93.5
69	f	12.67	152	60.5	105.0
70	f	15.92	191	63.3	113.5
71	f	15.83	190	66.8	140.0
72	f	11.67	140	60.0	77.0
73	f	12.33	148	60.5	84.5
74	f	15.75	189	64.3	113.5
75	f	11.92	143	58.3	77.5
76	f	14.83	178	66.5	117.5
77	f	13.67	164	65.3	98.0
78	f	13.08	157	60.5	112.0
79	f	12.25	147	59.5	101.0
80	f	12.33	148	59.0	95.0
81	f	14.75	177	61.3	81.0
82	f	14.25	171	61.5	91.0
83	f	14.33	172	64.8	142.0
84	f	15.83	190	56.8	98.5
85	f	15.25	183	66.5	112.0
86	f	11.92	143	61.5	116.5
87	f	14.92	179	63.0	98.5
88	f	15.50	186	57.0	83.5
89	f	15.17	182	65.5	133.0
90	f	15.17	182	62.0	91.5

```
 91   f   11.83   142   56.0    72.5
 92   f   13.75   165   61.3   106.5
 93   f   13.75   165   55.5    67.0
 94   f   12.83   154   61.0   122.5
 95   f   12.50   150   54.5    74.0
 96   f   12.92   155   66.0   144.5
 97   f   13.58   163   56.5    84.0
 98   f   11.75   141   56.0    72.5
 99   f   12.25   147   51.5    64.0
100   f   17.50   210   62.0   116.0
101   f   14.25   171   63.0    84.0
102   f   13.92   167   61.0    93.5
103   f   15.17   182   64.0   111.5
104   f   12.00   144   61.0    92.0
105   f   16.08   193   59.8   115.0
106   f   11.75   141   61.3    85.0
107   f   13.67   164   63.3   108.0
108   f   15.50   186   63.5   108.0
109   f   14.08   169   61.5    85.0
110   f   14.58   175   60.3    86.0
111   f   15.00   180   61.3   110.5
112   m   13.75   165   64.8    98.0
113   m   13.08   157   60.5   105.0
114   m   12.00   144   57.3    76.5
115   m   12.50   150   59.5    84.0
116   m   12.50   150   60.8   128.0
117   m   11.58   139   60.5    87.0
118   m   15.75   189   67.0   128.0
119   m   15.25   183   64.8   111.0
120   m   12.25   147   50.5    79.0
121   m   12.17   146   57.5    90.0
122   m   13.33   160   60.5    84.0
123   m   13.00   156   61.8   112.0
124   m   14.42   173   61.3    93.0
125   m   12.58   151   66.3   117.0
126   m   11.75   141   53.3    84.0
127   m   12.50   150   59.0    99.5
128   m   13.67   164   57.8    95.0
129   m   12.75   153   60.0    84.0
130   m   17.17   206   68.3   134.0
132   m   14.67   176   63.8    98.5
133   m   14.67   176   65.0   118.5
134   m   11.67   140   59.5    94.5
```

```
135    m    15.42    185    66.0    105.0
136    m    15.00    180    61.8    104.0
137    m    12.17    146    57.3     83.0
138    m    15.25    183    66.0    105.5
139    m    11.67    140    56.5     84.0
140    m    12.58    151    58.3     86.0
141    m    12.58    151    61.0     81.0
142    m    12.00    144    62.8     94.0
143    m    13.33    160    59.3     78.5
144    m    14.83    178    67.3    119.5
145    m    16.08    193    66.3    133.0
146    m    13.50    162    64.5    119.0
147    m    13.67    164    60.5     95.0
148    m    15.50    186    66.0    112.0
149    m    11.92    143    57.5     75.0
150    m    14.58    175    64.0     92.0
151    m    14.58    175    68.0    112.0
152    m    14.58    175    63.5     98.5
153    m    14.42    173    69.0    112.5
154    m    14.17    170    63.8    112.5
155    m    14.50    174    66.0    108.0
156    m    13.67    164    63.5    108.0
157    m    12.00    144    59.5     88.0
158    m    13.00    156    66.3    106.0
159    m    12.42    149    57.0     92.0
160    m    12.00    144    60.0    117.5
161    m    12.25    147    57.0     84.0
162    m    15.67    188    67.3    112.0
163    m    14.08    169    62.0    100.0
164    m    14.33    172    65.0    112.0
165    m    12.50    150    59.5     84.0
166    m    16.08    193    67.8    127.5
167    m    13.08    157    58.0     80.5
168    m    14.00    168    60.0     93.5
169    m    11.67    140    58.5     86.5
170    m    13.00    156    58.3     92.5
171    m    13.00    156    61.5    108.5
172    m    13.17    158    65.0    121.0
173    m    15.33    184    66.5    112.0
174    m    13.00    156    68.5    114.0
175    m    12.00    144    57.0     84.0
176    m    14.67    176    61.5     81.0
177    m    14.00    168    66.5    111.5
```

178	m	12.42	149	52.5	81.0
179	m	11.83	142	55.0	70.0
180	m	15.67	188	71.0	140.0
181	m	16.92	203	66.5	117.0
182	m	11.83	142	58.8	84.0
183	m	15.75	189	66.3	112.0
184	m	15.67	188	65.8	150.5
185	m	16.67	200	71.0	147.0
186	m	12.67	152	59.5	105.0
187	m	14.50	174	69.8	119.5
188	m	13.83	166	62.5	84.0
189	m	12.08	145	56.5	91.0
190	m	11.92	143	57.5	101.0
191	m	13.58	163	65.3	117.5
192	m	13.83	166	67.3	121.0
193	m	15.17	182	67.0	133.0
194	m	14.42	173	66.0	112.0
195	m	12.92	155	61.8	91.5
196	m	13.50	162	60.0	105.0
197	m	14.75	177	63.0	111.0
198	m	14.75	177	60.5	112.0
199	m	14.58	175	65.5	114.0
200	m	13.83	166	62.0	91.0
201	m	12.50	150	59.0	98.0
202	m	12.50	150	61.8	118.0
203	m	15.67	188	63.3	115.5
204	m	13.58	163	66.0	112.0
205	m	14.25	171	61.8	112.0
206	m	13.50	162	63.0	91.0
207	m	11.75	141	57.5	85.0
208	m	14.50	174	63.0	112.0
209	m	11.83	142	56.0	87.5
210	m	12.33	148	60.5	118.0
211	m	11.67	140	56.8	83.5
212	m	13.33	160	64.0	116.0
213	m	12.00	144	60.0	89.0
214	m	17.17	206	69.5	171.5
215	m	13.25	159	63.3	112.0
216	m	12.42	149	56.3	72.0
217	m	16.08	193	72.0	150.0
218	m	16.17	194	65.3	134.5
219	m	12.67	152	60.8	97.0
220	m	12.17	146	55.0	71.5

```
221    m    11.58    139    55.0    73.5
222    m    15.50    186    66.5   112.0
223    m    13.42    161    56.8    75.0
224    m    12.75    153    64.8   128.0
225    m    16.33    196    64.5    98.0
226    m    13.67    164    58.0    84.0
227    m    13.25    159    62.8    99.0
228    m    14.83    178    63.8   112.0
229    m    12.75    153    57.8    79.5
230    m    12.92    155    57.3    80.5
231    m    14.83    178    63.5   102.5
232    m    11.83    142    55.0    76.0
233    m    13.67    164    66.5   112.0
234    m    15.75    189    65.0   114.0
235    m    13.67    164    61.5   140.0
236    m    13.92    167    62.0   107.5
237    m    12.58    151    59.3    87.0
```

A quick preview of `heightweight` dataset.

```
ggplot(heightweight, aes(x = ageYear, y = heightIn)) +
  geom_point()
```

In this part of chart, `aes(x = ageYear, y = heightIn` specifies the aesthetics:

- `x = ageYear`: Assign `ageYear` as the x-axis label.
- `y = heightIn`: Assign `heightIn`as the y-axis label.

`geom_point()`:

- Adds multiple dots to the plots.
- `geom_point()` adds a layer of points to your plot, which creates a scatterplot.

```
labs(title = "Age and Height",
     x = 'Age (Year)',
     y = 'Height (Inch)',
     caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

To add more details on your chart, we can add `labs()` and `theme()` into our codes.

`labs()`:

- Used to customized axis labels and legend titles.
- To add `title` and `caption` unto the chart.

`theme()`:

- Used to give plots a consistent customized look.
- Positioning your chart `title` and adjusting the `size` of your chart font.

```
labs(title = "Age and Height by Gender",
     x = 'Age (Year)',
     y = 'Height (Inch)',
     caption = 'By Daven, DV, THU, 2024')
```

Inside `labs()`:

- `title = 'Histogram Chart'`: To add "Histogram Chart" as the `title`.
- `x = 'Age (Year)'`: Assign `'Age (Year)'` as the x-axis label.
- `y = 'Height (Inch)`: Assign `y = 'Height (Inch)` as the y-axis label.
- `caption = 'By Daven, DV, THU, 2024'`: To add "By Daven, DV, THU, 2024" as the `caption`.

```
theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Inside `theme()`:

- `plot.title = element_text()`: To customize the `title` chart to your desire output.

Inside `plot.title = element_text()`:

- `hjust = 0.5`: Positioning your chart `title` in the middle.
- `size = 20`: Sizing your title text to `size` 20.

```
library(gcookbook)
library(ggplot2)

ggplot(heightweight, aes(x = ageYear, y = heightIn)) +
  geom_point() +
  labs(title = "Age and Height",
       x = 'Age (Year)',
       y = 'Height (Inch)',
       caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Age and Height

By Daven, DV, THU, 2024

For a better view of the overall explanation above, this is the complete structure on how the codes *given above* should be arranged.

Note: You can use different size of your chart title and positioning (`hjust = 1, size = 15`).

20
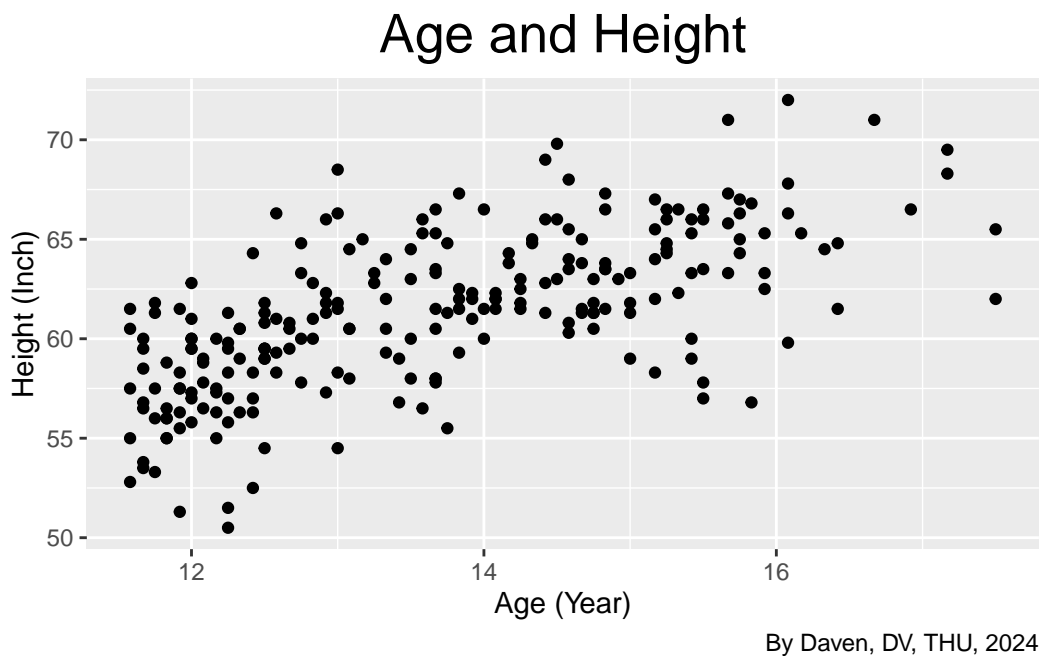
0.8 Correlation chart: Color by group

```
library(gcookbook)
library(ggplot2)
```

Using the same dataset in Correlation Chart we can expand further our bar chart by adding colors to the dots to add more visualization into it.

`heightweight`

```
   sex ageYear ageMonth heightIn weightLb
1    f   11.92      143     56.3     85.0
2    f   12.92      155     62.3    105.0
3    f   12.75      153     63.3    108.0
4    f   13.42      161     59.0     92.0
5    f   15.92      191     62.5    112.5
6    f   14.25      171     62.5    112.0
7    f   15.42      185     59.0    104.0
8    f   11.83      142     56.5     69.0
9    f   13.33      160     62.0     94.5
10   f   11.67      140     53.8     68.5
11   f   11.58      139     61.5    104.0
12   f   14.83      178     61.5    103.5
13   f   13.08      157     64.5    123.5
14   f   12.42      149     58.3     93.0
15   f   11.92      143     51.3     50.5
16   f   12.08      145     58.8     89.0
17   f   15.92      191     65.3    107.0
18   f   12.50      150     59.5     78.5
19   f   12.25      147     61.3    115.0
20   f   15.00      180     63.3    114.0
21   f   11.75      141     61.8     85.0
22   f   11.67      140     53.5     81.0
23   f   13.67      164     58.0     83.5
24   f   14.67      176     61.3    112.0
25   f   15.42      185     63.3    101.0
26   f   13.83      166     61.5    103.5
27   f   14.58      175     60.8     93.5
28   f   15.00      180     59.0    112.0
29   f   17.50      210     65.5    140.0
30   f   12.17      146     56.3     83.5
```

31	f	14.17	170	64.3	90.0
32	f	13.50	162	58.0	84.0
33	f	12.42	149	64.3	110.5
34	f	11.58	139	57.5	96.0
35	f	15.50	186	57.8	95.0
36	f	16.42	197	61.5	121.0
37	f	14.08	169	62.3	99.5
38	f	14.75	177	61.8	142.5
39	f	15.42	185	65.3	118.0
40	f	15.17	182	58.3	104.5
41	f	14.42	173	62.8	102.5
42	f	13.83	166	59.3	89.5
43	f	14.00	168	61.5	95.0
44	f	14.08	169	62.0	98.5
45	f	12.50	150	61.3	94.0
46	f	15.33	184	62.3	108.0
47	f	11.58	139	52.8	63.5
48	f	12.25	147	59.8	84.5
49	f	12.00	144	59.5	93.5
50	f	14.75	177	61.3	112.0
51	f	14.83	178	63.5	148.5
52	f	16.42	197	64.8	112.0
53	f	12.17	146	60.0	109.0
54	f	12.08	145	59.0	91.5
55	f	12.25	147	55.8	75.0
56	f	12.08	145	57.8	84.0
57	f	12.92	155	61.3	107.0
58	f	13.92	167	62.3	92.5
59	f	15.25	183	64.3	109.5
60	f	11.92	143	55.5	84.0
61	f	15.25	183	64.5	102.5
62	f	15.42	185	60.0	106.0
63	f	12.33	148	56.3	77.0
64	f	12.25	147	58.3	111.5
65	f	12.83	154	60.0	114.0
66	f	13.00	156	54.5	75.0
67	f	12.00	144	55.8	73.5
68	f	12.83	154	62.8	93.5
69	f	12.67	152	60.5	105.0
70	f	15.92	191	63.3	113.5
71	f	15.83	190	66.8	140.0
72	f	11.67	140	60.0	77.0
73	f	12.33	148	60.5	84.5

74	f	15.75	189	64.3	113.5
75	f	11.92	143	58.3	77.5
76	f	14.83	178	66.5	117.5
77	f	13.67	164	65.3	98.0
78	f	13.08	157	60.5	112.0
79	f	12.25	147	59.5	101.0
80	f	12.33	148	59.0	95.0
81	f	14.75	177	61.3	81.0
82	f	14.25	171	61.5	91.0
83	f	14.33	172	64.8	142.0
84	f	15.83	190	56.8	98.5
85	f	15.25	183	66.5	112.0
86	f	11.92	143	61.5	116.5
87	f	14.92	179	63.0	98.5
88	f	15.50	186	57.0	83.5
89	f	15.17	182	65.5	133.0
90	f	15.17	182	62.0	91.5
91	f	11.83	142	56.0	72.5
92	f	13.75	165	61.3	106.5
93	f	13.75	165	55.5	67.0
94	f	12.83	154	61.0	122.5
95	f	12.50	150	54.5	74.0
96	f	12.92	155	66.0	144.5
97	f	13.58	163	56.5	84.0
98	f	11.75	141	56.0	72.5
99	f	12.25	147	51.5	64.0
100	f	17.50	210	62.0	116.0
101	f	14.25	171	63.0	84.0
102	f	13.92	167	61.0	93.5
103	f	15.17	182	64.0	111.5
104	f	12.00	144	61.0	92.0
105	f	16.08	193	59.8	115.0
106	f	11.75	141	61.3	85.0
107	f	13.67	164	63.3	108.0
108	f	15.50	186	63.5	108.0
109	f	14.08	169	61.5	85.0
110	f	14.58	175	60.3	86.0
111	f	15.00	180	61.3	110.5
112	m	13.75	165	64.8	98.0
113	m	13.08	157	60.5	105.0
114	m	12.00	144	57.3	76.5
115	m	12.50	150	59.5	84.0
116	m	12.50	150	60.8	128.0

117	m	11.58	139	60.5	87.0
118	m	15.75	189	67.0	128.0
119	m	15.25	183	64.8	111.0
120	m	12.25	147	50.5	79.0
121	m	12.17	146	57.5	90.0
122	m	13.33	160	60.5	84.0
123	m	13.00	156	61.8	112.0
124	m	14.42	173	61.3	93.0
125	m	12.58	151	66.3	117.0
126	m	11.75	141	53.3	84.0
127	m	12.50	150	59.0	99.5
128	m	13.67	164	57.8	95.0
129	m	12.75	153	60.0	84.0
130	m	17.17	206	68.3	134.0
132	m	14.67	176	63.8	98.5
133	m	14.67	176	65.0	118.5
134	m	11.67	140	59.5	94.5
135	m	15.42	185	66.0	105.0
136	m	15.00	180	61.8	104.0
137	m	12.17	146	57.3	83.0
138	m	15.25	183	66.0	105.5
139	m	11.67	140	56.5	84.0
140	m	12.58	151	58.3	86.0
141	m	12.58	151	61.0	81.0
142	m	12.00	144	62.8	94.0
143	m	13.33	160	59.3	78.5
144	m	14.83	178	67.3	119.5
145	m	16.08	193	66.3	133.0
146	m	13.50	162	64.5	119.0
147	m	13.67	164	60.5	95.0
148	m	15.50	186	66.0	112.0
149	m	11.92	143	57.5	75.0
150	m	14.58	175	64.0	92.0
151	m	14.58	175	68.0	112.0
152	m	14.58	175	63.5	98.5
153	m	14.42	173	69.0	112.5
154	m	14.17	170	63.8	112.5
155	m	14.50	174	66.0	108.0
156	m	13.67	164	63.5	108.0
157	m	12.00	144	59.5	88.0
158	m	13.00	156	66.3	106.0
159	m	12.42	149	57.0	92.0
160	m	12.00	144	60.0	117.5

161	m	12.25	147	57.0	84.0
162	m	15.67	188	67.3	112.0
163	m	14.08	169	62.0	100.0
164	m	14.33	172	65.0	112.0
165	m	12.50	150	59.5	84.0
166	m	16.08	193	67.8	127.5
167	m	13.08	157	58.0	80.5
168	m	14.00	168	60.0	93.5
169	m	11.67	140	58.5	86.5
170	m	13.00	156	58.3	92.5
171	m	13.00	156	61.5	108.5
172	m	13.17	158	65.0	121.0
173	m	15.33	184	66.5	112.0
174	m	13.00	156	68.5	114.0
175	m	12.00	144	57.0	84.0
176	m	14.67	176	61.5	81.0
177	m	14.00	168	66.5	111.5
178	m	12.42	149	52.5	81.0
179	m	11.83	142	55.0	70.0
180	m	15.67	188	71.0	140.0
181	m	16.92	203	66.5	117.0
182	m	11.83	142	58.8	84.0
183	m	15.75	189	66.3	112.0
184	m	15.67	188	65.8	150.5
185	m	16.67	200	71.0	147.0
186	m	12.67	152	59.5	105.0
187	m	14.50	174	69.8	119.5
188	m	13.83	166	62.5	84.0
189	m	12.08	145	56.5	91.0
190	m	11.92	143	57.5	101.0
191	m	13.58	163	65.3	117.5
192	m	13.83	166	67.3	121.0
193	m	15.17	182	67.0	133.0
194	m	14.42	173	66.0	112.0
195	m	12.92	155	61.8	91.5
196	m	13.50	162	60.0	105.0
197	m	14.75	177	63.0	111.0
198	m	14.75	177	60.5	112.0
199	m	14.58	175	65.5	114.0
200	m	13.83	166	62.0	91.0
201	m	12.50	150	59.0	98.0
202	m	12.50	150	61.8	118.0
203	m	15.67	188	63.3	115.5

```
204     m    13.58     163     66.0     112.0
205     m    14.25     171     61.8     112.0
206     m    13.50     162     63.0      91.0
207     m    11.75     141     57.5      85.0
208     m    14.50     174     63.0     112.0
209     m    11.83     142     56.0      87.5
210     m    12.33     148     60.5     118.0
211     m    11.67     140     56.8      83.5
212     m    13.33     160     64.0     116.0
213     m    12.00     144     60.0      89.0
214     m    17.17     206     69.5     171.5
215     m    13.25     159     63.3     112.0
216     m    12.42     149     56.3      72.0
217     m    16.08     193     72.0     150.0
218     m    16.17     194     65.3     134.5
219     m    12.67     152     60.8      97.0
220     m    12.17     146     55.0      71.5
221     m    11.58     139     55.0      73.5
222     m    15.50     186     66.5     112.0
223     m    13.42     161     56.8      75.0
224     m    12.75     153     64.8     128.0
225     m    16.33     196     64.5      98.0
226     m    13.67     164     58.0      84.0
227     m    13.25     159     62.8      99.0
228     m    14.83     178     63.8     112.0
229     m    12.75     153     57.8      79.5
230     m    12.92     155     57.3      80.5
231     m    14.83     178     63.5     102.5
232     m    11.83     142     55.0      76.0
233     m    13.67     164     66.5     112.0
234     m    15.75     189     65.0     114.0
235     m    13.67     164     61.5     140.0
236     m    13.92     167     62.0     107.5
237     m    12.58     151     59.3      87.0
```

A quick preview of `heightweight` dataset.

```
ggplot(heightweight, aes(x = ageYear, y = heightIn, color = sex)) +
  geom_point()
```

In this part of chart, `aes(x = ageYear, y = heightIn, color = sex)` specifies the aesthetics:

- `x = ageYear`: Assign `ageYear` as the x-axis label.
- `y = heightIn`: Assign `heightIn`as the y-axis label.
- `color = sex`: Assign `color` to `sex` (male, female).

`geom_point()`:

- Adds multiple dots to the plots.
- `geom_point()` adds a layer of points to your plot, which creates a scatterplot.

```
labs(title = "Age and Height by Gender",
      x = 'Age (Year)',
      y = 'Height (Inch)',
      caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

To add more details on your chart, we can add `labs()` and `theme()` into our codes.

`labs()`:

- Used to customized axis labels and legend titles.
- To add `title` and `caption` unto the chart.

```
theme():
```

- Used to give plots a consistent customized look.
- Positioning your chart `title` and adjusting the `size` of your chart font.

```
labs(title = "Age and Height by Gender",
      x = 'Age (Year)',
      y = 'Height (Inch)',
      caption = 'By Daven, DV, THU, 2024')
```

Inside `labs()`:

- `title = 'Histogram Chart'`: To add "Histogram Chart" as the `title`.
- `x = 'Age (Year)'`: Assign `'Age (Year)'` as the x-axis label.
- `y = 'Height (Inch)`: Assign `y = 'Height (Inch)` as the y-axis label.
- `caption = 'By Daven, DV, THU, 2024'`: To add "By Daven, DV, THU, 2024" as the `caption`.

```
theme(plot.title = element_text(hjust = 0.5, size = 20))
```

```
List of 1
 $ plot.title:List of 11
  ..$ family       : NULL
  ..$ face         : NULL
  ..$ colour       : NULL
  ..$ size         : num 20
  ..$ hjust        : num 0.5
  ..$ vjust        : NULL
  ..$ angle        : NULL
  ..$ lineheight   : NULL
  ..$ margin       : NULL
  ..$ debug        : NULL
  ..$ inherit.blank: logi FALSE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 - attr(*, "class")= chr [1:2] "theme" "gg"
 - attr(*, "complete")= logi FALSE
 - attr(*, "validate")= logi TRUE
```

Inside `theme()`:

- `plot.title = element_text()`: To customize the `title` chart to your desire output.

Inside `plot.title = element_text()`:

- `hjust = 0.5`: Positioning your chart `title` in the middle.
- `size = 20`: Sizing your title text to `size` 20.

```r
library(gcookbook)
library(ggplot2)

ggplot(heightweight, aes(x = ageYear, y = heightIn, color = sex)) +
  geom_point() +
  labs(title = "Age and Height by Gender",
       x = 'Age (Year)',
       y = 'Height (Inch)',
       caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```
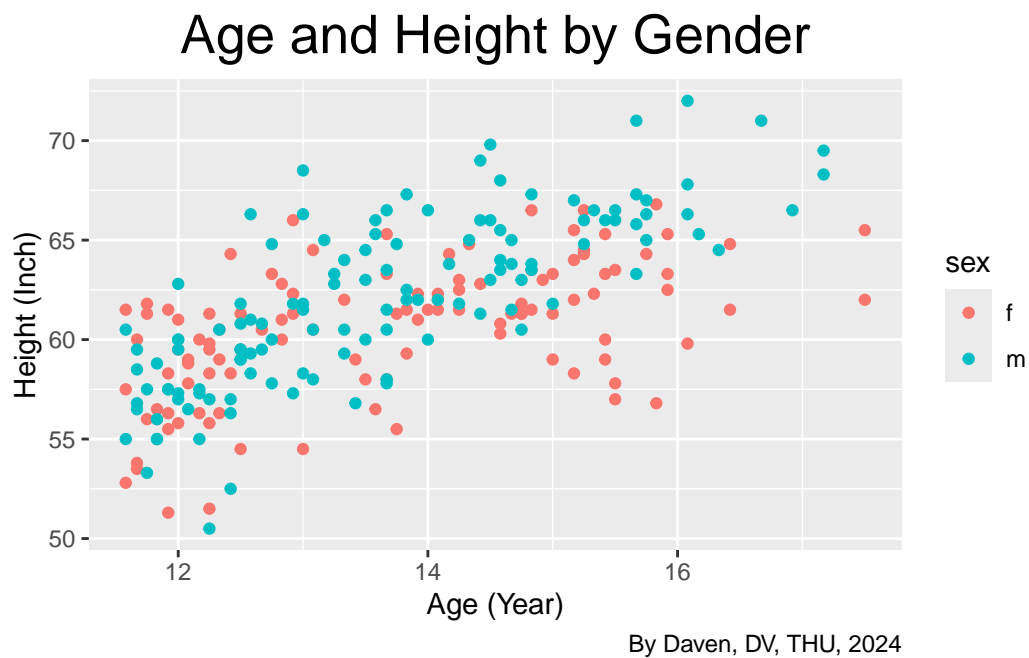

For a better view of the overall explanation above, this is the complete structure on how the codes *given above* should be arranged.

Note: You can use different size of your chart title and positioning (`hjust = 1, size = 15`).

0.9 Multigroup histogram

Similar to Histogram, in this part we'll be combining 2 different `group` into 1 singular chart.

```
library(ggplot2)
library(MASS)
library(tidyverse)
```

Here, we need to initialize the `birthwt` dataset by using the `MASS` library and `tidyverse` library to combine the 2 separate chart into 1 chart.

```
birthwt
```

	low	age	lwt	race	smoke	ptl	ht	ui	ftv	bwt
85	0	19	182	2	0	0	0	1	0	2523
86	0	33	155	3	0	0	0	0	3	2551
87	0	20	105	1	1	0	0	0	1	2557
88	0	21	108	1	1	0	0	1	2	2594
89	0	18	107	1	1	0	0	1	0	2600
91	0	21	124	3	0	0	0	0	0	2622
92	0	22	118	1	0	0	0	0	1	2637
93	0	17	103	3	0	0	0	0	1	2637
94	0	29	123	1	1	0	0	0	1	2663
95	0	26	113	1	1	0	0	0	0	2665
96	0	19	95	3	0	0	0	0	0	2722
97	0	19	150	3	0	0	0	0	1	2733
98	0	22	95	3	0	0	1	0	0	2751
99	0	30	107	3	0	1	0	1	2	2750
100	0	18	100	1	1	0	0	0	0	2769
101	0	18	100	1	1	0	0	0	0	2769
102	0	15	98	2	0	0	0	0	0	2778
103	0	25	118	1	1	0	0	0	3	2782
104	0	20	120	3	0	0	0	1	0	2807
105	0	28	120	1	1	0	0	0	1	2821
106	0	32	121	3	0	0	0	0	2	2835
107	0	31	100	1	0	0	0	1	3	2835
108	0	36	202	1	0	0	0	0	1	2836
109	0	28	120	3	0	0	0	0	0	2863
111	0	25	120	3	0	0	0	1	2	2877
112	0	28	167	1	0	0	0	0	0	2877
113	0	17	122	1	1	0	0	0	0	2906
114	0	29	150	1	0	0	0	0	2	2920

115	0	26	168	2	1	0	0	0	0	2920
116	0	17	113	2	0	0	0	0	1	2920
117	0	17	113	2	0	0	0	0	1	2920
118	0	24	90	1	1	1	0	0	1	2948
119	0	35	121	2	1	1	0	0	1	2948
120	0	25	155	1	0	0	0	0	1	2977
121	0	25	125	2	0	0	0	0	0	2977
123	0	29	140	1	1	0	0	0	2	2977
124	0	19	138	1	1	0	0	0	2	2977
125	0	27	124	1	1	0	0	0	0	2922
126	0	31	215	1	1	0	0	0	2	3005
127	0	33	109	1	1	0	0	0	1	3033
128	0	21	185	2	1	0	0	0	2	3042
129	0	19	189	1	0	0	0	0	2	3062
130	0	23	130	2	0	0	0	0	1	3062
131	0	21	160	1	0	0	0	0	0	3062
132	0	18	90	1	1	0	0	1	0	3062
133	0	18	90	1	1	0	0	1	0	3062
134	0	32	132	1	0	0	0	0	4	3080
135	0	19	132	3	0	0	0	0	0	3090
136	0	24	115	1	0	0	0	0	2	3090
137	0	22	85	3	1	0	0	0	0	3090
138	0	22	120	1	0	0	1	0	1	3100
139	0	23	128	3	0	0	0	0	0	3104
140	0	22	130	1	1	0	0	0	0	3132
141	0	30	95	1	1	0	0	0	2	3147
142	0	19	115	3	0	0	0	0	0	3175
143	0	16	110	3	0	0	0	0	0	3175
144	0	21	110	3	1	0	0	1	0	3203
145	0	30	153	3	0	0	0	0	0	3203
146	0	20	103	3	0	0	0	0	0	3203
147	0	17	119	3	0	0	0	0	0	3225
148	0	17	119	3	0	0	0	0	0	3225
149	0	23	119	3	0	0	0	0	2	3232
150	0	24	110	3	0	0	0	0	0	3232
151	0	28	140	1	0	0	0	0	0	3234
154	0	26	133	3	1	2	0	0	0	3260
155	0	20	169	3	0	1	0	1	1	3274
156	0	24	115	3	0	0	0	0	2	3274
159	0	28	250	3	1	0	0	0	6	3303
160	0	20	141	1	0	2	0	1	1	3317
161	0	22	158	2	0	1	0	0	2	3317
162	0	22	112	1	1	2	0	0	0	3317

163	0	31	150	3	1	0	0	0	2	3321
164	0	23	115	3	1	0	0	0	1	3331
166	0	16	112	2	0	0	0	0	0	3374
167	0	16	135	1	1	0	0	0	0	3374
168	0	18	229	2	0	0	0	0	0	3402
169	0	25	140	1	0	0	0	0	1	3416
170	0	32	134	1	1	1	0	0	4	3430
172	0	20	121	2	1	0	0	0	0	3444
173	0	23	190	1	0	0	0	0	0	3459
174	0	22	131	1	0	0	0	0	1	3460
175	0	32	170	1	0	0	0	0	0	3473
176	0	30	110	3	0	0	0	0	0	3544
177	0	20	127	3	0	0	0	0	0	3487
179	0	23	123	3	0	0	0	0	0	3544
180	0	17	120	3	1	0	0	0	0	3572
181	0	19	105	3	0	0	0	0	0	3572
182	0	23	130	1	0	0	0	0	0	3586
183	0	36	175	1	0	0	0	0	0	3600
184	0	22	125	1	0	0	0	0	1	3614
185	0	24	133	1	0	0	0	0	0	3614
186	0	21	134	3	0	0	0	0	2	3629
187	0	19	235	1	1	0	1	0	0	3629
188	0	25	95	1	1	3	0	1	0	3637
189	0	16	135	1	1	0	0	0	0	3643
190	0	29	135	1	0	0	0	0	1	3651
191	0	29	154	1	0	0	0	0	1	3651
192	0	19	147	1	1	0	0	0	0	3651
193	0	19	147	1	1	0	0	0	0	3651
195	0	30	137	1	0	0	0	0	1	3699
196	0	24	110	1	0	0	0	0	1	3728
197	0	19	184	1	1	0	1	0	0	3756
199	0	24	110	3	0	1	0	0	0	3770
200	0	23	110	1	0	0	0	0	1	3770
201	0	20	120	3	0	0	0	0	0	3770
202	0	25	241	2	0	0	1	0	0	3790
203	0	30	112	1	0	0	0	0	1	3799
204	0	22	169	1	0	0	0	0	0	3827
205	0	18	120	1	1	0	0	0	2	3856
206	0	16	170	2	0	0	0	0	4	3860
207	0	32	186	1	0	0	0	0	2	3860
208	0	18	120	3	0	0	0	0	1	3884
209	0	29	130	1	1	0	0	0	2	3884
210	0	33	117	1	0	0	0	1	1	3912

211	0	20	170	1	1	0	0	0	0 3940
212	0	28	134	3	0	0	0	0	1 3941
213	0	14	135	1	0	0	0	0	0 3941
214	0	28	130	3	0	0	0	0	0 3969
215	0	25	120	1	0	0	0	0	2 3983
216	0	16	95	3	0	0	0	0	1 3997
217	0	20	158	1	0	0	0	0	1 3997
218	0	26	160	3	0	0	0	0	0 4054
219	0	21	115	1	0	0	0	0	1 4054
220	0	22	129	1	0	0	0	0	0 4111
221	0	25	130	1	0	0	0	0	2 4153
222	0	31	120	1	0	0	0	0	2 4167
223	0	35	170	1	0	1	0	0	1 4174
224	0	19	120	1	1	0	0	0	0 4238
225	0	24	116	1	0	0	0	0	1 4593
226	0	45	123	1	0	0	0	0	1 4990
4	1	28	120	3	1	1	0	1	0 709
10	1	29	130	1	0	0	0	1	2 1021
11	1	34	187	2	1	0	1	0	0 1135
13	1	25	105	3	0	1	1	0	0 1330
15	1	25	85	3	0	0	0	1	0 1474
16	1	27	150	3	0	0	0	0	0 1588
17	1	23	97	3	0	0	0	1	1 1588
18	1	24	128	2	0	1	0	0	1 1701
19	1	24	132	3	0	0	1	0	0 1729
20	1	21	165	1	1	0	1	0	1 1790
22	1	32	105	1	1	0	0	0	0 1818
23	1	19	91	1	1	2	0	1	0 1885
24	1	25	115	3	0	0	0	0	0 1893
25	1	16	130	3	0	0	0	0	1 1899
26	1	25	92	1	1	0	0	0	0 1928
27	1	20	150	1	1	0	0	0	2 1928
28	1	21	200	2	0	0	0	1	2 1928
29	1	24	155	1	1	1	0	0	0 1936
30	1	21	103	3	0	0	0	0	0 1970
31	1	20	125	3	0	0	0	1	0 2055
32	1	25	89	3	0	2	0	0	1 2055
33	1	19	102	1	0	0	0	0	2 2082
34	1	19	112	1	1	0	0	1	0 2084
35	1	26	117	1	1	1	0	0	0 2084
36	1	24	138	1	0	0	0	0	0 2100
37	1	17	130	3	1	1	0	1	0 2125
40	1	20	120	2	1	0	0	0	3 2126

```
42    1   22 130     1      1    1   0   1    1 2187
43    1   27 130     2      0    0   0   1    0 2187
44    1   20  80     3      1    0   0   1    0 2211
45    1   17 110     1      1    0   0   0    0 2225
46    1   25 105     3      0    1   0   0    1 2240
47    1   20 109     3      0    0   0   0    0 2240
49    1   18 148     3      0    0   0   0    0 2282
50    1   18 110     2      1    1   0   0    0 2296
51    1   20 121     1      1    1   0   1    0 2296
52    1   21 100     3      0    1   0   0    4 2301
54    1   26  96     3      0    0   0   0    0 2325
56    1   31 102     1      1    1   0   0    1 2353
57    1   15 110     1      0    0   0   0    0 2353
59    1   23 187     2      1    0   0   0    1 2367
60    1   20 122     2      1    0   0   0    0 2381
61    1   24 105     2      1    0   0   0    0 2381
62    1   15 115     3      0    0   0   1    0 2381
63    1   23 120     3      0    0   0   0    0 2410
65    1   30 142     1      1    1   0   0    0 2410
67    1   22 130     1      1    0   0   0    1 2410
68    1   17 120     1      1    0   0   0    3 2414
69    1   23 110     1      1    1   0   0    0 2424
71    1   17 120     2      0    0   0   0    2 2438
75    1   26 154     3      0    1   1   0    1 2442
76    1   20 105     3      0    0   0   0    3 2450
77    1   26 190     1      1    0   0   0    0 2466
78    1   14 101     3      1    1   0   0    0 2466
79    1   28  95     1      1    0   0   0    2 2466
81    1   14 100     3      0    0   0   0    2 2495
82    1   23  94     3      1    0   0   0    0 2495
83    1   17 142     2      0    0   1   0    0 2495
84    1   21 130     1      1    0   1   0    3 2495
```

A quick preview of the `birthwt` dataset.

```
birthwt_mod <- birthwt

birthwt_mod$smoke <- recode_factor(birthwt_mod$smoke, '0' = 'No Smoke', '1' = 'Smoke')
```

Before proceeding we need to make another copy the dataset make some modification into it.

```
birthwt_mod <- birthwt:
```

- This to make a copy of the original dataset, so any modification will go to `birthwt_mod`.

`recode_factor`:

- Converts or modifies the `smoke` variable to a factor (categorical variable).
- Making the values are interpreted as categories instead of numeric.

`birthwt_mod$smoke, '0' = 'No Smoke', '1' = 'Smoke':`

- Values in `smoke` that are 0 are replaced with the label `'No Smoke'`.
- Values in `smoke` that are 1 are replaced with the label `'Smoke'`.

```
ggplot(birthwt_mod, aes(x = bwt, fill = smoke)) +
  geom_histogram(position = "identity", alpha = 0.4)
```

In this part of chart, `birthwt_mod, aes(x = bwt, fill = smoke)` specifies the aesthetics:

- `x = bwt`: Assign `bwt` variable to the x-axis.
- `fill = smoke`: Assign color based on the `smoke` variable ('No Smoke' and 'Smoke').

`geom_histogram`:

- Adds a frequency of column geometry similar to a bar to the plot.
- `geom_histogram()` visualizes the distribution of a continuous variable by dividing it into bins.

`position = "identity", alpha = 0.4`:

- `position = "identity"`: Overlays the bars for both groups ('No Smoke' and 'Smoke').
- `alpha = 0.4`: Sets the transparency level for the bars to `0.4`

```
labs(title = "Birth Weight",
     x = 'Birth Weight',
     y = 'Count',
     caption = 'By By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

To add more details on your chart, we can add `labs()` and `theme()` into our codes.

`labs()`:

- Used to customized axis labels and legend titles.
- To add `title` and `caption` unto the chart.

`theme()`:

- Used to give plots a consistent customized look.
- Positioning your chart `title` and adjusting the `size` of your chart font.

```
labs(title = "Birth Weight",
     x = 'Birth Weight',
     y = 'Count',
     caption = 'By By Daven, DV, THU, 2024')
```

Inside `labs()`:

- `title = 'Birth Weight'`: To add "Birth Weight" as the `title`.
- `x = 'Birth Weight'`: Assign `'Birth Weight'` as the x-axis label.
- `y = 'Count'`: Assign `'Count'` as the y-axis label.
- `caption = 'By Daven, DV, THU, 2024'`: To add "By Daven, DV, THU, 2024" as the `caption`.

```
theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Inside `theme()`:

- `plot.title = element_text()`: To customize the `title` chart to your desire output.

Inside `plot.title = element_text()`:

- `hjust = 0.5`: Positioning your chart `title` in the middle.
- `size = 20`: Sizing your title text to `size` 20.

```
library(ggplot2)
library(MASS)
library(tidyverse)

birthwt_mod <- birthwt

birthwt_mod$smoke <- recode_factor(birthwt_mod$smoke, '0' = 'No Smoke', '1' = 'Smoke')

ggplot(birthwt_mod, aes(x = bwt, fill = smoke)) +
  geom_histogram(position = "identity", alpha = 0.4) +
  labs(title = "Birth Weight",
       x = 'Birth Weight',
       y = 'Count',
       caption = 'By By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Birth Weight
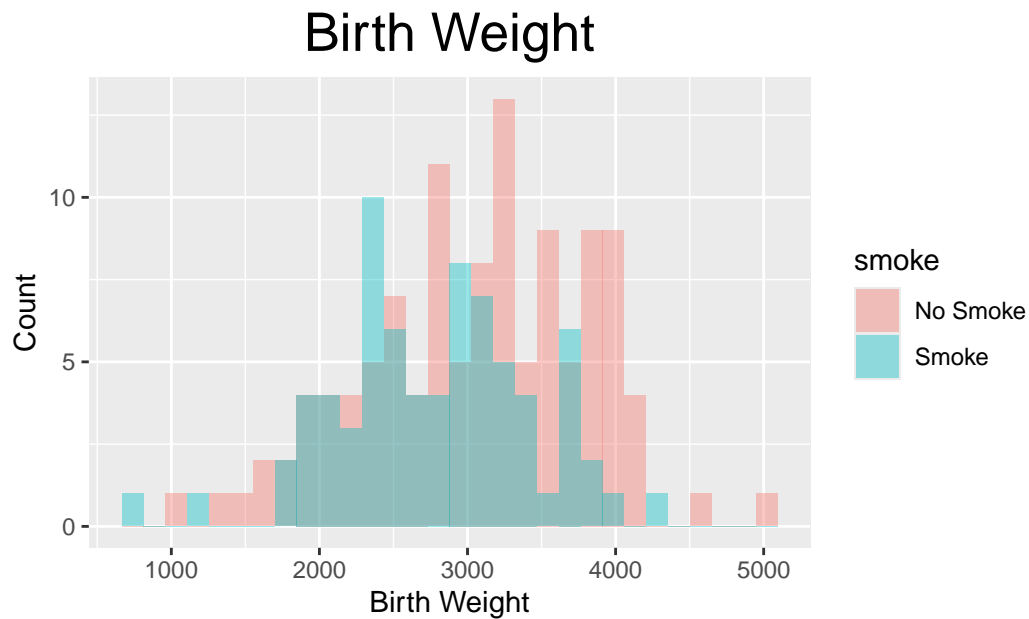
For a better view of the overall explanation above, this is the complete structure on how the codes *given above* should be arranged.

Note: You can use different size of your chart title and positioning (`hjust = 1, size = 15`).

0.10 Density chart

In this section, will be plotting a density chart using `faithful` dataset.

```
library(ggplot2)
library(tidyr)
```

To start, we'll need to use the `tidyr` library to use the `faithful` dataset.

```
faithful
```

```
  eruptions waiting
1     3.600      79
2     1.800      54
3     3.333      74
4     2.283      62
```

5	4.533	85
6	2.883	55
7	4.700	88
8	3.600	85
9	1.950	51
10	4.350	85
11	1.833	54
12	3.917	84
13	4.200	78
14	1.750	47
15	4.700	83
16	2.167	52
17	1.750	62
18	4.800	84
19	1.600	52
20	4.250	79
21	1.800	51
22	1.750	47
23	3.450	78
24	3.067	69
25	4.533	74
26	3.600	83
27	1.967	55
28	4.083	76
29	3.850	78
30	4.433	79
31	4.300	73
32	4.467	77
33	3.367	66
34	4.033	80
35	3.833	74
36	2.017	52
37	1.867	48
38	4.833	80
39	1.833	59
40	4.783	90
41	4.350	80
42	1.883	58
43	4.567	84
44	1.750	58
45	4.533	73
46	3.317	83
47	3.833	64

48	2.100	53
49	4.633	82
50	2.000	59
51	4.800	75
52	4.716	90
53	1.833	54
54	4.833	80
55	1.733	54
56	4.883	83
57	3.717	71
58	1.667	64
59	4.567	77
60	4.317	81
61	2.233	59
62	4.500	84
63	1.750	48
64	4.800	82
65	1.817	60
66	4.400	92
67	4.167	78
68	4.700	78
69	2.067	65
70	4.700	73
71	4.033	82
72	1.967	56
73	4.500	79
74	4.000	71
75	1.983	62
76	5.067	76
77	2.017	60
78	4.567	78
79	3.883	76
80	3.600	83
81	4.133	75
82	4.333	82
83	4.100	70
84	2.633	65
85	4.067	73
86	4.933	88
87	3.950	76
88	4.517	80
89	2.167	48
90	4.000	86

91	2.200	60
92	4.333	90
93	1.867	50
94	4.817	78
95	1.833	63
96	4.300	72
97	4.667	84
98	3.750	75
99	1.867	51
100	4.900	82
101	2.483	62
102	4.367	88
103	2.100	49
104	4.500	83
105	4.050	81
106	1.867	47
107	4.700	84
108	1.783	52
109	4.850	86
110	3.683	81
111	4.733	75
112	2.300	59
113	4.900	89
114	4.417	79
115	1.700	59
116	4.633	81
117	2.317	50
118	4.600	85
119	1.817	59
120	4.417	87
121	2.617	53
122	4.067	69
123	4.250	77
124	1.967	56
125	4.600	88
126	3.767	81
127	1.917	45
128	4.500	82
129	2.267	55
130	4.650	90
131	1.867	45
132	4.167	83
133	2.800	56

134	4.333	89
135	1.833	46
136	4.383	82
137	1.883	51
138	4.933	86
139	2.033	53
140	3.733	79
141	4.233	81
142	2.233	60
143	4.533	82
144	4.817	77
145	4.333	76
146	1.983	59
147	4.633	80
148	2.017	49
149	5.100	96
150	1.800	53
151	5.033	77
152	4.000	77
153	2.400	65
154	4.600	81
155	3.567	71
156	4.000	70
157	4.500	81
158	4.083	93
159	1.800	53
160	3.967	89
161	2.200	45
162	4.150	86
163	2.000	58
164	3.833	78
165	3.500	66
166	4.583	76
167	2.367	63
168	5.000	88
169	1.933	52
170	4.617	93
171	1.917	49
172	2.083	57
173	4.583	77
174	3.333	68
175	4.167	81
176	4.333	81

177	4.500	73
178	2.417	50
179	4.000	85
180	4.167	74
181	1.883	55
182	4.583	77
183	4.250	83
184	3.767	83
185	2.033	51
186	4.433	78
187	4.083	84
188	1.833	46
189	4.417	83
190	2.183	55
191	4.800	81
192	1.833	57
193	4.800	76
194	4.100	84
195	3.966	77
196	4.233	81
197	3.500	87
198	4.366	77
199	2.250	51
200	4.667	78
201	2.100	60
202	4.350	82
203	4.133	91
204	1.867	53
205	4.600	78
206	1.783	46
207	4.367	77
208	3.850	84
209	1.933	49
210	4.500	83
211	2.383	71
212	4.700	80
213	1.867	49
214	3.833	75
215	3.417	64
216	4.233	76
217	2.400	53
218	4.800	94
219	2.000	55

220	4.150	76
221	1.867	50
222	4.267	82
223	1.750	54
224	4.483	75
225	4.000	78
226	4.117	79
227	4.083	78
228	4.267	78
229	3.917	70
230	4.550	79
231	4.083	70
232	2.417	54
233	4.183	86
234	2.217	50
235	4.450	90
236	1.883	54
237	1.850	54
238	4.283	77
239	3.950	79
240	2.333	64
241	4.150	75
242	2.350	47
243	4.933	86
244	2.900	63
245	4.583	85
246	3.833	82
247	2.083	57
248	4.367	82
249	2.133	67
250	4.350	74
251	2.200	54
252	4.450	83
253	3.567	73
254	4.500	73
255	4.150	88
256	3.817	80
257	3.917	71
258	4.450	83
259	2.000	56
260	4.283	79
261	4.767	78
262	4.533	84

```
263     1.850       58
264     4.250       83
265     1.983       43
266     2.250       60
267     4.750       75
268     4.117       81
269     2.150       46
270     4.417       90
271     1.817       46
272     4.467       74
```

A quick preview of `faithful` dataset.

```
ggplot(faithful, aes(x = waiting, y = ..density..)) +
  geom_density(fill = "blue", alpha = .2) +
  xlim(35, 105)
```
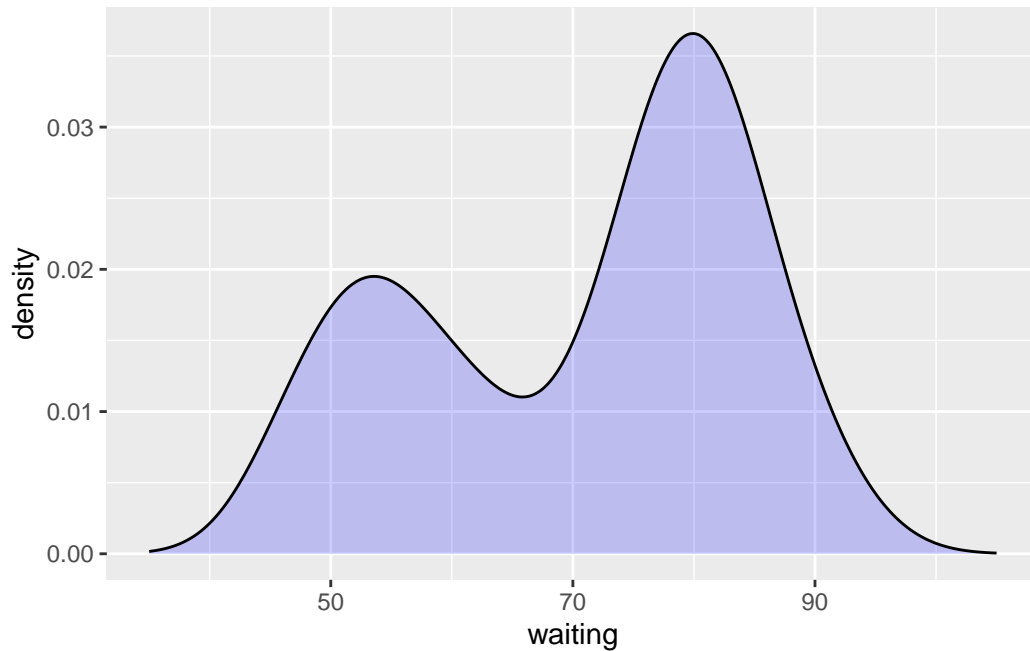

In this part of chart, `aes(x = waiting, y = ..density..)` specifies the aesthetics:

- `x = waiting`: Assign `waiting` variable to the x-axis.
- `y = ..density..`: Assign `..density..` variable to the y-axis. (`..density..` a special variable computed by `geom_density()`).

`geom_density`:

- Adds a density plot layer.
- `geom_density` creates smooth continuous variable version of the histogram plot.

Inside `geom_density` command, `fill = "blue", alpha = .2` specifies the following:

- `fill = "blue"`: Assign color to the bar's area using the color `"blue"`.
- `alpha = .2`: Sets the transparency level for the bars to `.2`.

`xlim(35, 105)`:

- Sets the limits of the x-axis to the range `(35, 105)`.
- Values outside this range are excluded from the plot.

```
abs(title = "Density Chart",
      x = 'Waiting',
      y = 'Density',
      caption = 'By By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

To add more details on your chart, we can add `labs()` and `theme()` into our codes.

`labs()`:

- Used to customized axis labels and legend titles.
- To add `title` and `caption` unto the chart.

`theme()`:

- Used to give plots a consistent customized look.
- Positioning your chart `title` and adjusting the `size` of your chart font.

```
labs(title = 'Density Chart',
      x = 'Waiting',
      y = 'Density',
      caption = 'By By Daven, DV, THU, 2024')
```

Inside `labs()`:

- `title = 'Density Chart'`: To add "Density Chart" as the `title`.
- `x = 'Waiting'`: Assign `'Waiting'` as the x-axis label.
- `y = 'Density'`: Assign `'Density'` as the y-axis label.
- `caption = 'By Daven, DV, THU, 2024'`: To add "By Daven, DV, THU, 2024" as the `caption`.

```
theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Inside `theme()`:

- `plot.title = element_text()`: To customize the `title` chart to your desire output.

Inside `plot.title = element_text()`:

- `hjust = 0.5`: Positioning your chart `title` in the middle.
- `size = 20`: Sizing your title text to `size` 20.

```
library(ggplot2)
library(tidyr)

ggplot(faithful, aes(x = waiting, y = ..density..)) +
  geom_density(fill = "blue", alpha = .2) +
  xlim(35, 105) +
  labs(title = "Density Chart",
       x = 'Waiting',
       y = 'Density',
       caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```
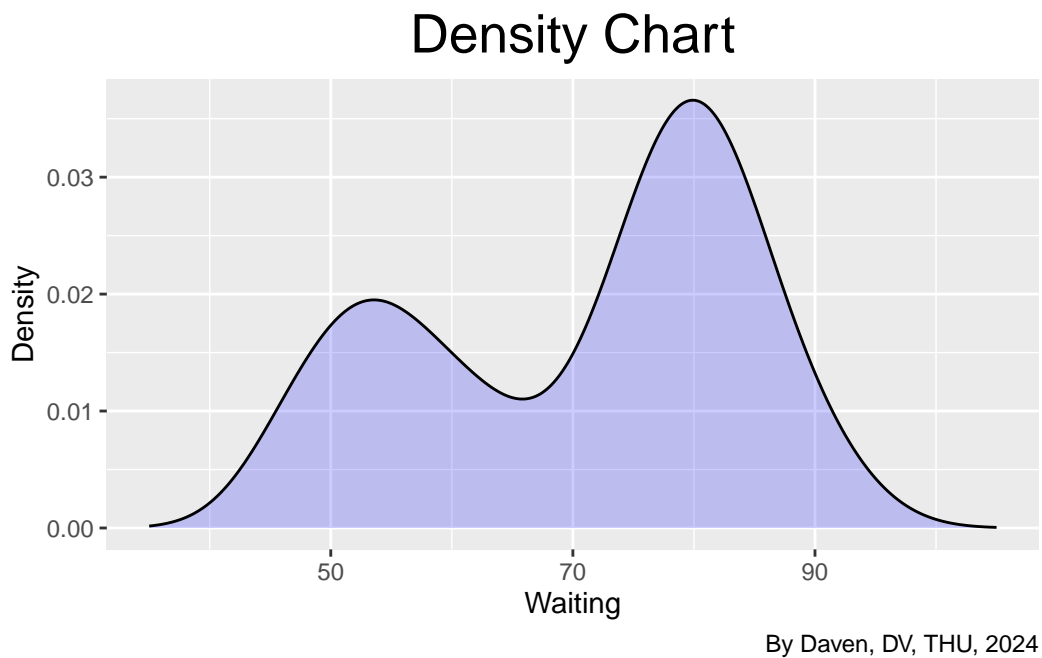
Density Chart

By Daven, DV, THU, 2024

For a better view of the overall explanation above, this is the complete structure on how the codes *given above* should be arranged.

Note: You can use different colors on chart's curve like `'red'` *or* `'green'` *as well as the size of your chart title and positioning (*`hjust = 1, size = 15`*).*

0.11 Box plot

Our last section, will be plotting `ToothGrowth` dataset using a boxplot chart.

```
library(gcookbook)
library(ggplot2)
```

Here we need to initializes `ToothGrowth` dataset by using `gcookbook()` library within our code.

```
ToothGrowth
```

```
    len supp dose
1    4.2   VC  0.5
2   11.5   VC  0.5
3    7.3   VC  0.5
4    5.8   VC  0.5
5    6.4   VC  0.5
6   10.0   VC  0.5
7   11.2   VC  0.5
8   11.2   VC  0.5
9    5.2   VC  0.5
10   7.0   VC  0.5
11  16.5   VC  1.0
12  16.5   VC  1.0
13  15.2   VC  1.0
14  17.3   VC  1.0
15  22.5   VC  1.0
16  17.3   VC  1.0
17  13.6   VC  1.0
18  14.5   VC  1.0
19  18.8   VC  1.0
20  15.5   VC  1.0
21  23.6   VC  2.0
22  18.5   VC  2.0
23  33.9   VC  2.0
```

```
24 25.5   VC  2.0
25 26.4   VC  2.0
26 32.5   VC  2.0
27 26.7   VC  2.0
28 21.5   VC  2.0
29 23.3   VC  2.0
30 29.5   VC  2.0
31 15.2   OJ  0.5
32 21.5   OJ  0.5
33 17.6   OJ  0.5
34  9.7   OJ  0.5
35 14.5   OJ  0.5
36 10.0   OJ  0.5
37  8.2   OJ  0.5
38  9.4   OJ  0.5
39 16.5   OJ  0.5
40  9.7   OJ  0.5
41 19.7   OJ  1.0
42 23.3   OJ  1.0
43 23.6   OJ  1.0
44 26.4   OJ  1.0
45 20.0   OJ  1.0
46 25.2   OJ  1.0
47 25.8   OJ  1.0
48 21.2   OJ  1.0
49 14.5   OJ  1.0
50 27.3   OJ  1.0
51 25.5   OJ  2.0
52 26.4   OJ  2.0
53 22.4   OJ  2.0
54 24.5   OJ  2.0
55 24.8   OJ  2.0
56 30.9   OJ  2.0
57 26.4   OJ  2.0
58 27.3   OJ  2.0
59 29.4   OJ  2.0
60 23.0   OJ  2.0
```

A quick preview of `ToothGrowth` dataset.

```
ggplot(ToothGrowth, aes(x = interaction(supp, dose), y = len)) +
  geom_boxplot()
```
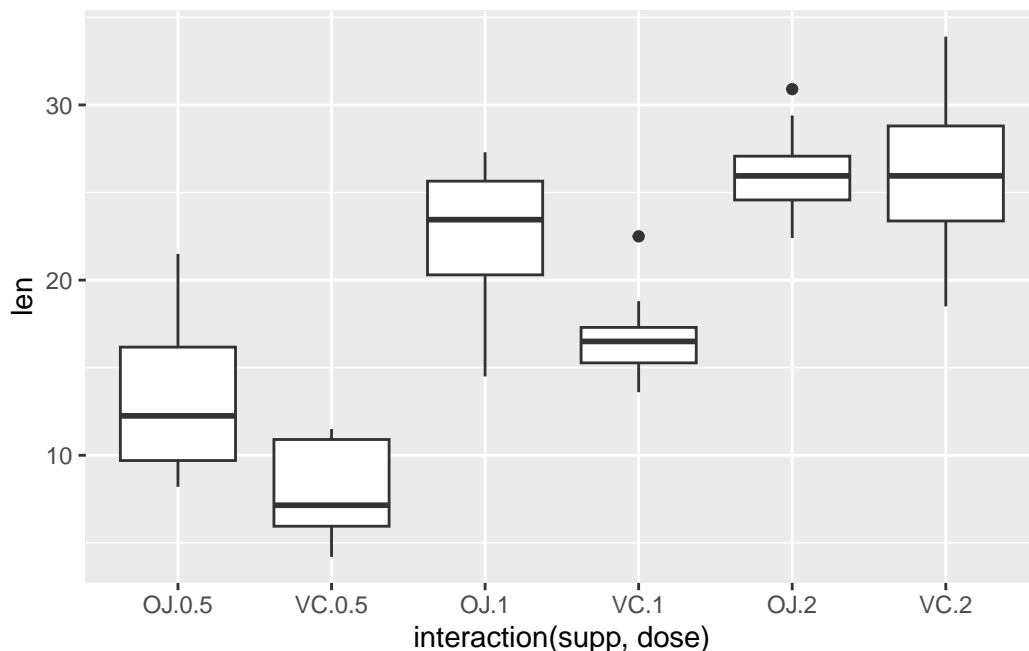
In this part of chart, `aes(x = interaction(supp, dose), y = len)` specifies the aesthetics:

- `x = interaction(supp, dose)`: Assign `interaction(supp, dose)` variable to the x-axis label by combining levels of `supp` (VC or OJ) and `dose` (0.5, 1, or 2).
- `y = len`: Assign `len` variable to the y-axis.

`geom_boxplot()`:

- Adds a boxplot layer to the plot.
- `geom_boxplot` summarizes the distribution of `len` for each group defined by the interaction of `supp` and `dose`.

```
labs(title = "Boxplot with ggplot",
     x = 'Supp + Dose',
     y = 'Length',
     fill = "Smoking Status",
     caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```

To add more details on your chart, we can add `labs()` and `theme()` into our codes.

`labs()`:

- Used to customized axis labels and legend titles.

- To add `title` and `caption` unto the chart.

`theme()`:

- Used to give plots a consistent customized look.
- Positioning your chart `title` and adjusting the `size` of your chart font.

```
labs(title = 'Boxplot with ggplot',
     x = 'Supp + Dose',
     y = 'Length',
     fill = 'Smoking Status',
     caption = 'By Daven, DV, THU, 2024')
```

Inside `labs()`:

- `title = 'Boxplot with ggplot'`: To add "Boxplot with ggplot" as the `title`.
- `x = 'Supp + Dose'`: Assign `'Supp + Dose'` as the x-axis label.
- `y = 'Length'`: Assign `'Length'` as the y-axis label.
- `fill = 'Smoking Status'`: Assign the legend `title` for the `fill` to `'Smoking Status'`.
- `caption = 'By Daven, DV, THU, 2024'`: To add "By Daven, DV, THU, 2024" as the `caption`.

```
theme(plot.title = element_text(hjust = 0.5, size = 20))
```

Inside `theme()`:

- `plot.title = element_text()`: To customize the `title` chart to your desire output.
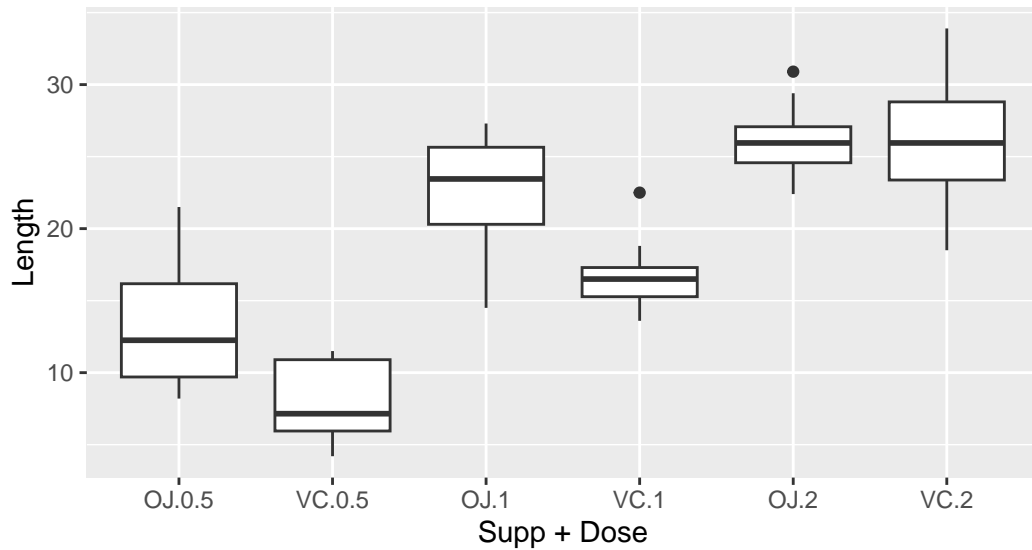
Inside `plot.title = element_text()`:

- `hjust = 0.5`: Positioning your chart `title` in the middle.
- `size = 20`: Sizing your title text to `size` 20.

```
library(gcookbook)
library(ggplot2)

ggplot(ToothGrowth, aes(x = interaction(supp, dose), y = len)) +
  geom_boxplot() +
  labs(title = "Boxplot with ggplot",
       x = 'Supp + Dose',
       y = 'Length',
       fill = "Smoking Status",
```

```
      caption = 'By Daven, DV, THU, 2024') +
  theme(plot.title = element_text(hjust = 0.5, size = 20))
```


Boxplot with ggplot

By Daven, DV, THU, 2024

For a better view of the overall explanation above, this is the complete structure on how the codes *given above* should be arranged.

Note: You can use different size of your chart title and positioning (`hjust = 1, size = 15`).