Exploring data #1

Geoms

Adding geoms

There are a number of different `geom_*` functions you can use to add geoms to a plot. They are divided between geoms that directly map the data to an aesthetic and those that show some summary or statistic of the data.

Some of the most common "statistical" geoms are:

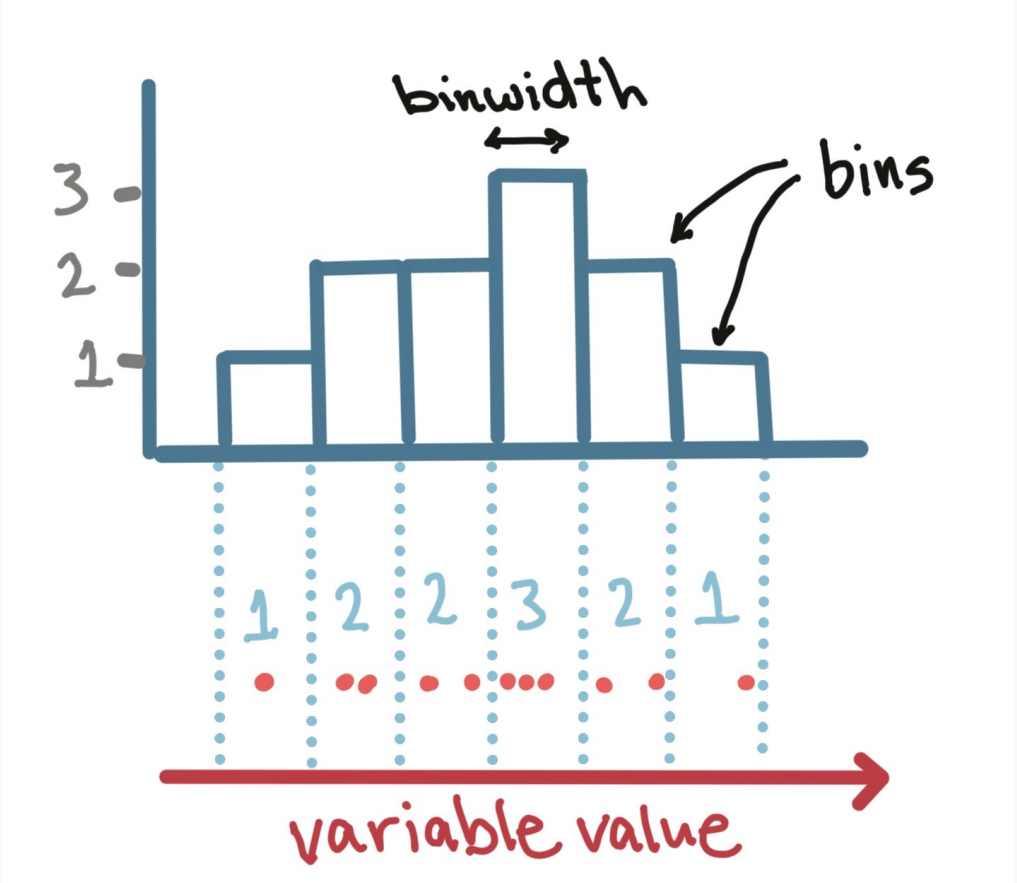
Geom(s)	Description
`geom_histogram`	Show distribution in 1-D
`geom_hex`, `geom_density`	Show distribution in 2-D
`geom_col`, `geom_bar`	Create bar charts
`geom_boxplot`, `geom_dotplot`	Create boxplots and related plots
`geom_smooth`	Add a fitted line to a scatterplot

Adding geoms

These “statistical” geoms all input the original data and perform some calculations on that data to determine how to plot the final geom. Often, this calculation involves some kind of summarization.

Adding geoms

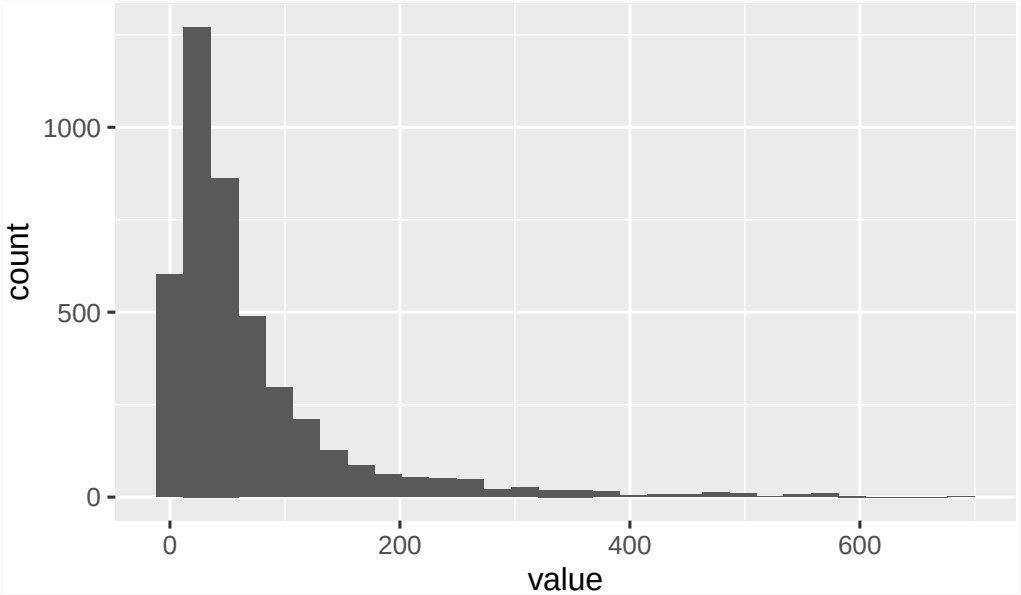
For example, the geom for a histogram (`geom_hist`) divides the data into an evenly-sized set of "bins" and then calculates the number of points in each bin to provide a visualization of how the data is distributed.



Adding geoms

To plot a histogram of PM2.5 concentrations in the Beijing data, run:

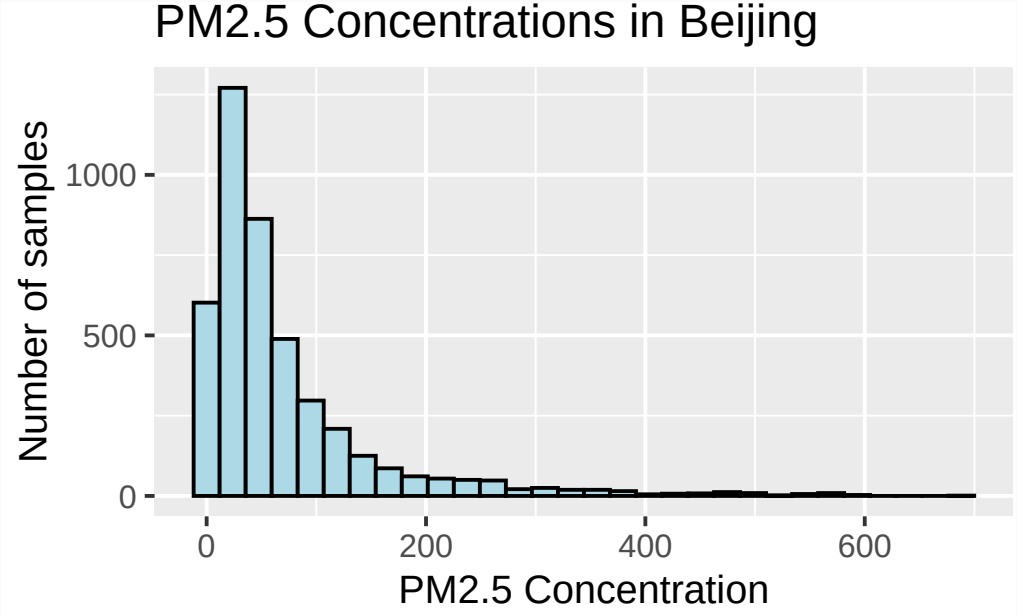
```
ggplot(data = beijing_pm) +
  geom_histogram(aes(x = value))
```



Histogram example

You can add some elements to the histogram, like `ggtitle`, and `labs`:

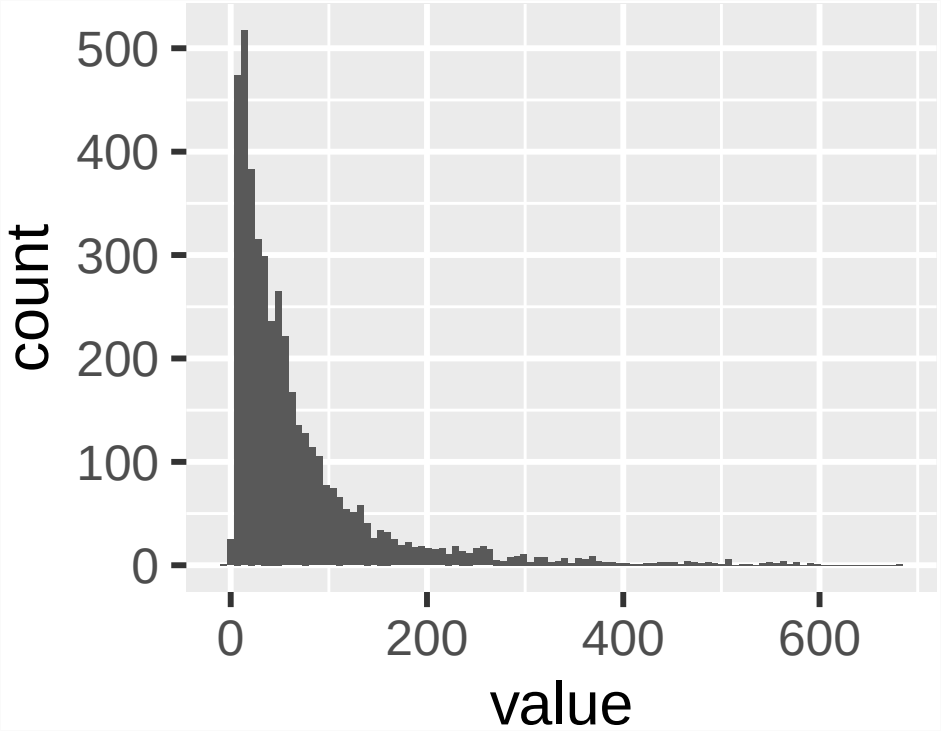
```
ggplot(beijing_pm, aes(x = value)) +
  geom_histogram(fill = "lightblue", color = "black") +
  ggtitle("PM2.5 Concentrations in Beijing") +
  labs(x = "PM2.5 Concentration", y = "Number of samples")
```



Histogram example

`geom_histogram` also has its own special argument, `bins`. You can use this to change the number of bins that are used to make the histogram:

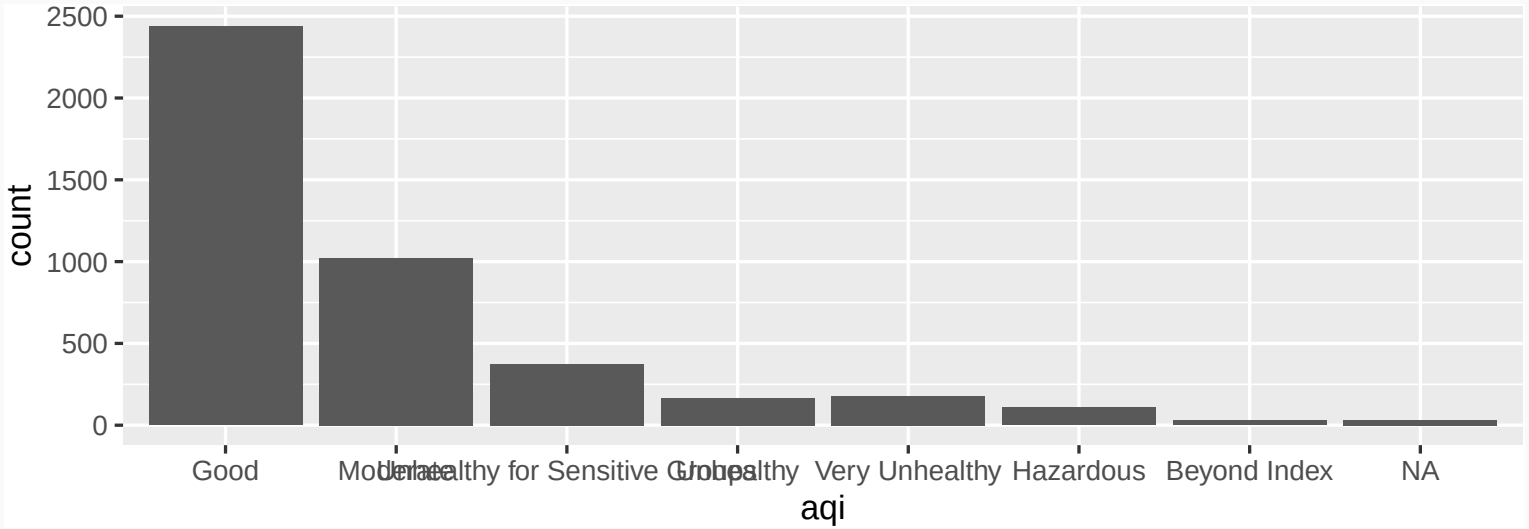
```
ggplot(beijing_pm, aes(x = value)) +
  geom_histogram(bins = 100)
```



Bar chart

You can use the `geom_bar` geom to create a barchart:

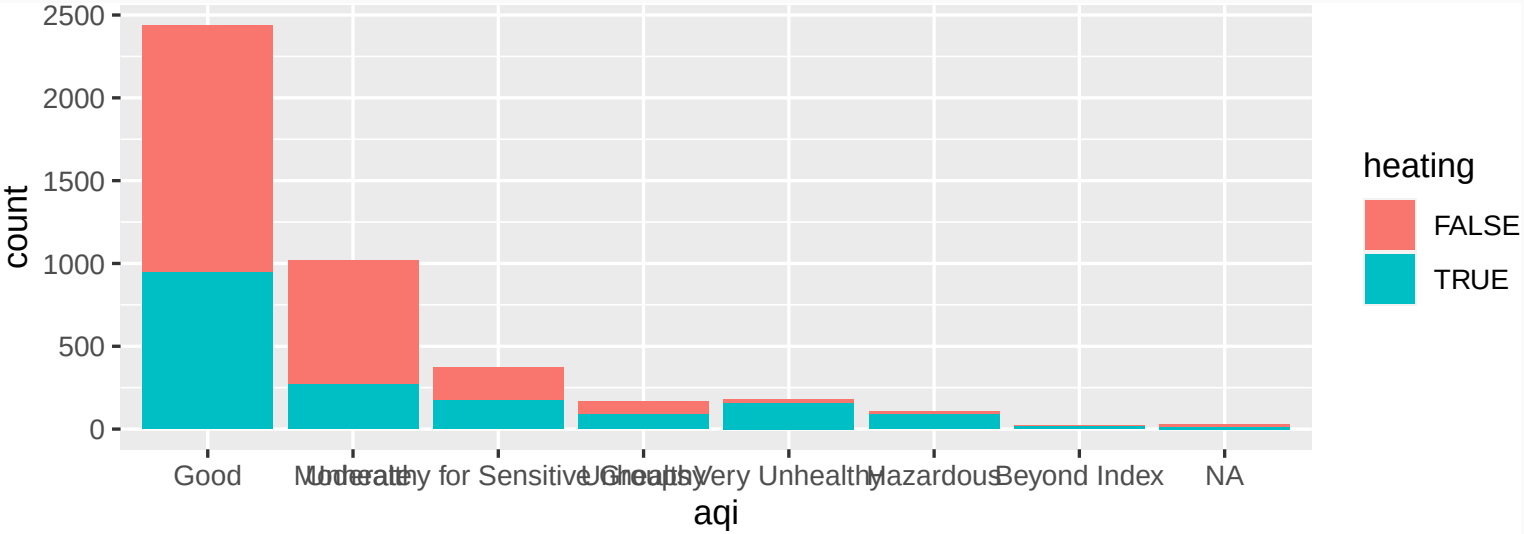
```
ggplot(beijing_pm, aes(x = aqi)) +
  geom_bar()
```



Bar chart

You can use the `geom_bar` geom to show counts for two factors by using `x` for one and `fill` for the other:

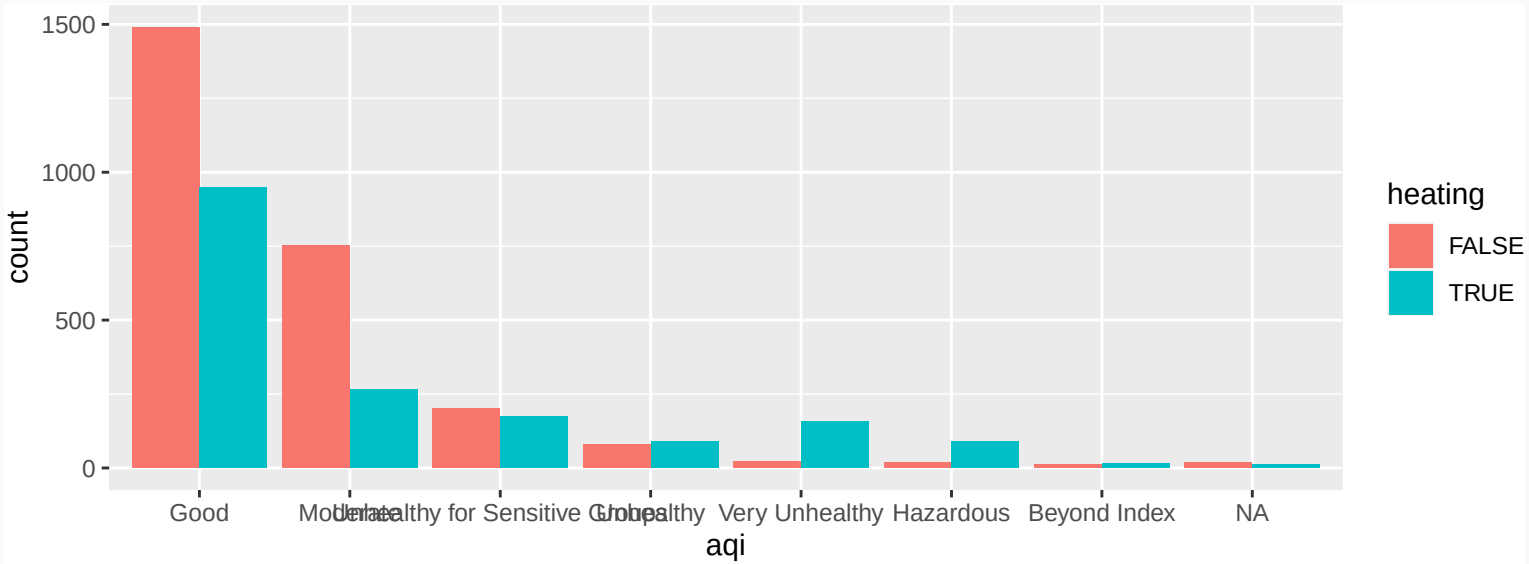
```
ggplot(beijing_pm, aes(x = aqi, fill = heating)) +
  geom_bar()
```



Bar chart

With the `geom_bar` geom, you can use the `position` argument to change how the bars for different groups are shown (`"stack"`, `"dodge"`, `"fill"`):

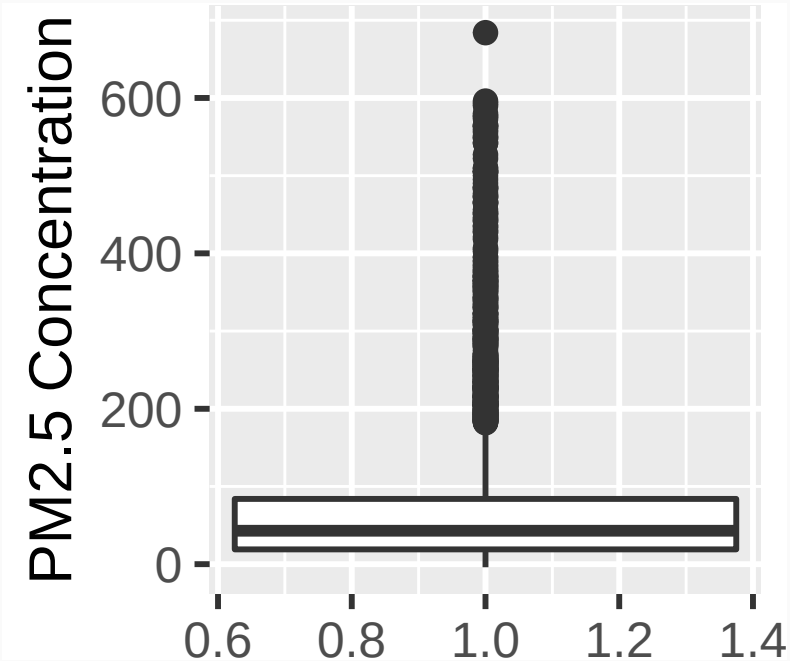
```
ggplot(beijing_pm, aes(x = aqi, fill = heating)) +
  geom_bar(position = "dodge")
```



Boxplot example

To create a boxplot, you can use `geom_boxplot`:

```
ggplot(beijing_pm, aes(x = 1, y = value)) +
  geom_boxplot() +
  labs(x = "", y = "PM2.5 Concentration")
```



Boxplot example

You can also do separate boxplots by a factor. In this case, you'll need to include two aesthetics (`x` and `y`) when you initialize the ggplot object.

```
ggplot(beijing_pm, aes(x = aqi, y = value, group = aqi)) +
  geom_boxplot() +
  labs(x = "AQI Category", y = "PM2.5 Concentration")
```