

Q-Q Plots

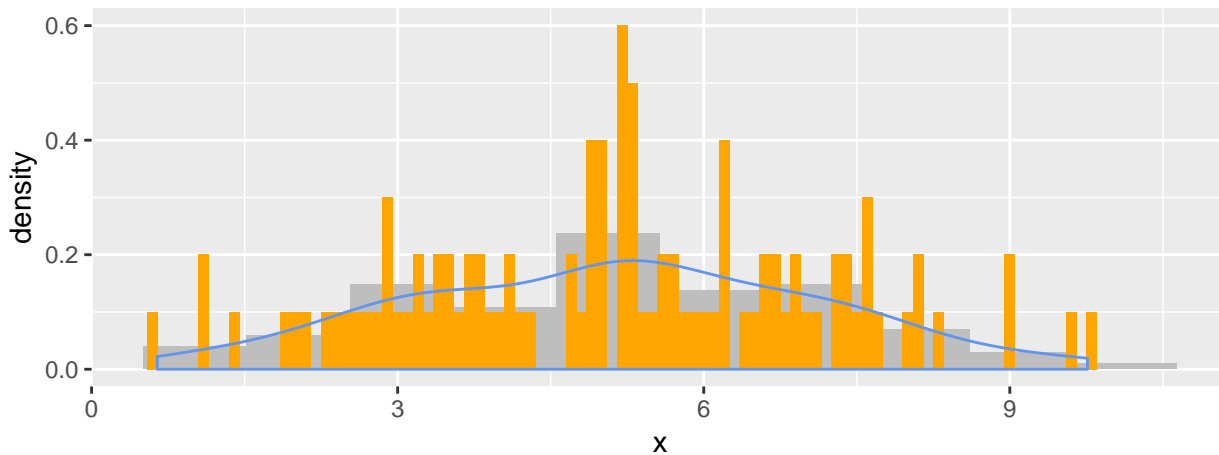
```
library(dplyr)
library(ggplot2)
library(gridExtra)
```

Setting: We have n observations of a random variable, and we want to know if they follow a normal distribution.

```
set.seed(9789426)
example_data <- data.frame(
  x = rnorm(100, 5, 2)
)
```

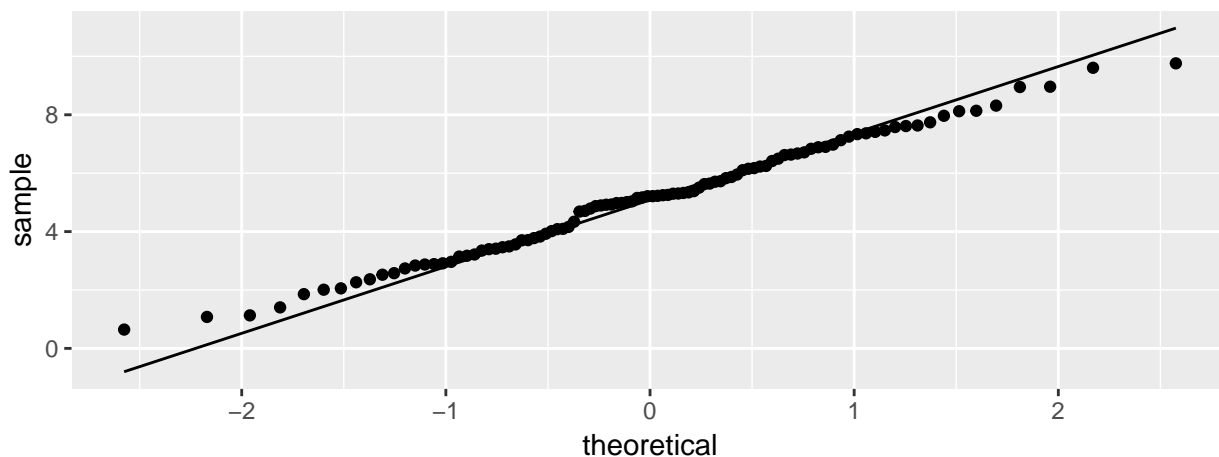
One approach, good enough for most purposes: a density plot or histogram.

```
ggplot(data = example_data, mapping = aes(x = x)) +
  geom_histogram(bins = 10, fill = "gray", mapping = aes(y = ..density..)) +
  geom_histogram(binwidth = 0.1, fill = "orange", mapping = aes(y = ..density..)) +
  geom_density(color = "cornflowerblue")
```



Another option is the Quantile-Quantile plot. Here we compare our sample to a standard Normal(0, 1) distribution (the default):

```
ggplot(data = example_data, mapping = aes(sample = x)) +
  stat_qq() +
  stat_qq_line()
```



Each point in this plot is of the form

(q -th quantile of theoretical distribution, q -th quantile of sample)

For example, if we pick $q = 0.1$, we can find a point in the plot above as follows:

The x axis coordinate is the value x such that $P(Z \leq x) = 0.1$, where $Z \sim N(0, 1)$:

```
qnorm(0.1, mean = 0, sd = 1)
```

```
## [1] -1.281552
```

The y axis coordinate is the value y such that 10 percent of our sample data are less than y :

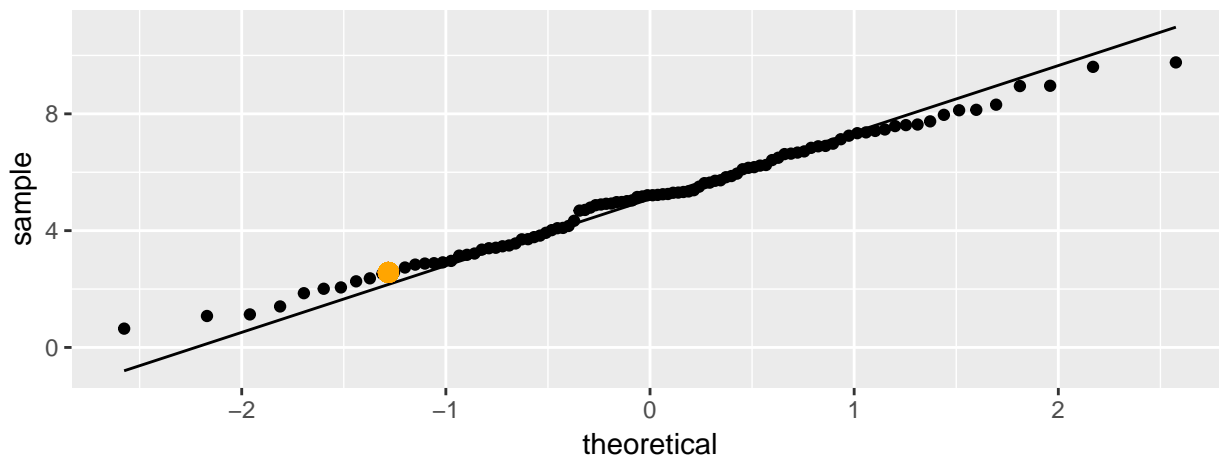
```
quantile(example_data$x, 0.1)
```

```
##      10%
```

```
## 2.570362
```

So, the point $(-1.281552, 2.570362)$ is in our Q-Q plot:

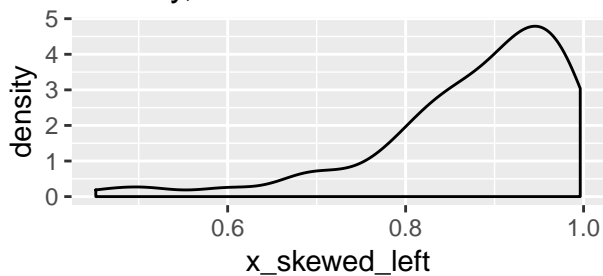
```
ggplot(data = example_data, mapping = aes(sample = x)) +  
  stat_qq() +  
  stat_qq_line() +  
  geom_point(x = -1.281552, y = 2.570362, color = "orange", size = 3)
```



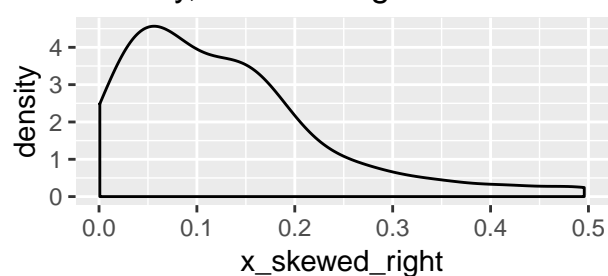
If the sample data come from a normal distribution, the points should fall close to a line: there is an approximately linear relationship between sample and theoretical quantiles.

If the data come from some other distribution, we will see some non-linear pattern in the plot. Here are some examples based on samples from skewed distributions.

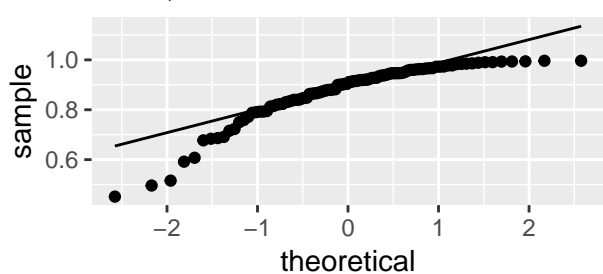
Density, Skewed Left



Density, Skewed Right



Q-Q, Skewed Left



Q-Q, Skewed Right

