

Friday the 13th

```
library(readr)
library(ggplot2)
library(dplyr)
```

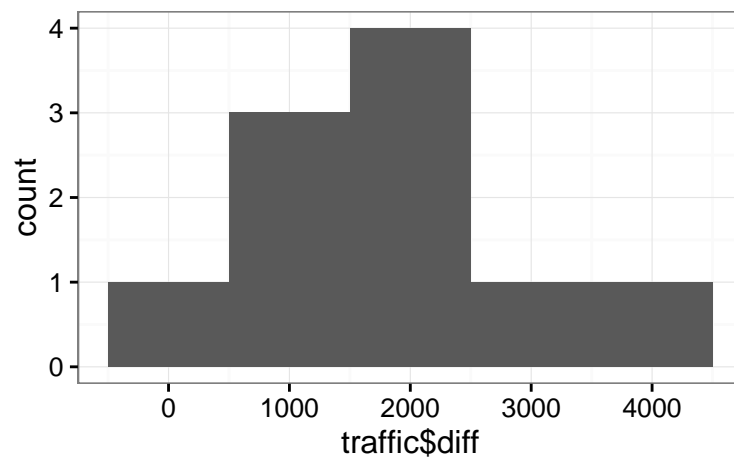
```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
theme_set(theme_bw())
```

Read in the data

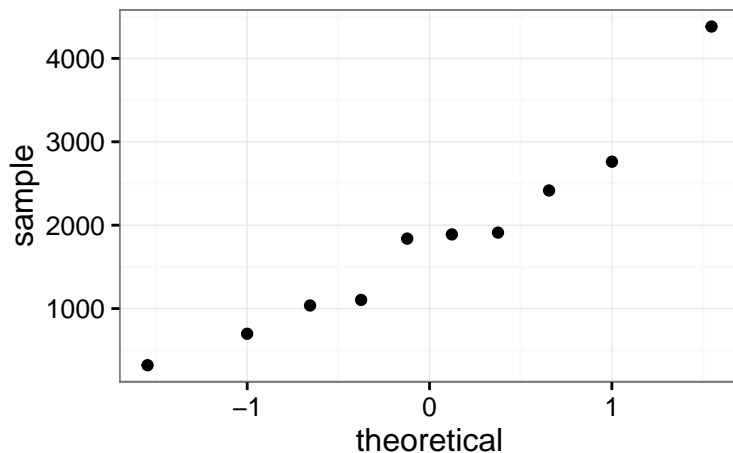
```
friday <- read.delim("https://dyurovsky.github.io/psyc20100/data/demos/friday.txt")
traffic <- friday %>%
  filter(type == "traffic")
```

Let's look at some basic descriptives

```
#Plot a histogram
qplot(traffic$diff, binwidth = 1000)
```



```
# Make a qqplot
ggplot(traffic, aes(samples = diff)) +
  stat_qq()
```



```
# Descriptive statistics
descriptives <- traffic %>%
  summarise(mean = mean(diff),
            sd = sd(diff),
            n = n())
```

```
descriptives
```

```
##      mean      sd  n
## 1 1835.8 1176.014 10
```

Ok let's compute t-test statistics by hand

```
# Get the components I need for the T-statistic formula
test_statistics <- descriptives %>%
  mutate(df = n - 1,
         se = sd/sqrt(n),
         t = (mean - 0)/se)
```

```
test_statistics
```

```
##      mean      sd  n df      se      t
## 1 1835.8 1176.014 10  9 371.8882 4.936429
```

```
# Find the p-value using position in the t-distribution with the appropriate degrees of freedom
p_val <- 2 * pt(test_statistics$t, test_statistics$df, lower.tail = FALSE)
p_val
```

```
## [1] 0.0008061844
```

Ok now let's skip all of that and use the built-in R function

```
# Use the t.test function to ske
friday_test <- t.test(traffic$diff, alternative = "two.sided")
friday_test
```

```
##
## One Sample t-test
##
## data: traffic$diff
## t = 4.9364, df = 9, p-value = 0.0008062
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
```

```
## 994.5304 2677.0696
## sample estimates:
## mean of x
## 1835.8
```

Make a plot of our empirical data on the Null Hypothesis distribution

```
t_9 <- data.frame(x = seq(-6,6,.01),
                  y = dt(seq(-6,6,.01), 9))

ggplot(data = t_9, aes(x = x, y = y)) +
  geom_point(size = .25) + #plot the null distribution
  geom_vline(aes(xintercept = test_statistics$t), size = 2, color = "darkred") + #empirical data
  geom_vline(aes(xintercept = qt(.025, 9)), color = "gray", size = 1.5) + #2.5th percentile
  geom_vline(aes(xintercept = qt(.975, 9)), color = "gray", size = 1.5) + #97.5th percentile
  theme(panel.grid = element_blank()) +
  ylab("") +
  xlab("t-value")
```

