# Data Visualization

*Swati Jani Joshi*

*June 30, 2016*

## Titanic

R Markdown containig the code for Data Visualization Titanic exercise.

## Load Data

```r
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
```

```r
library(tidyr)
library(ggplot2)
titanic <- read.csv("~/Documents/Data Wrangling/DW_2/titanic_original.csv")
```

## Structure of titanic

```r
str(titanic)
```

```
## 'data.frame':    1309 obs. of  14 variables:
##  $ pclass   : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ survived : int  1 1 0 0 0 1 1 0 1 0 ...
##  $ name     : Factor w/ 1307 levels "Abbing, Mr. Anthony",..: 22 24 25 26 27 31 46 47 51 55 ...
##  $ sex      : Factor w/ 2 levels "female","male": 1 2 1 2 1 2 1 2 1 2 ...
##  $ age      : num  29 0.917 2 30 25 ...
##  $ sibsp    : int  0 1 1 1 1 0 1 0 2 0 ...
##  $ parch    : int  0 2 2 2 2 0 0 0 0 0 ...
##  $ ticket   : Factor w/ 929 levels "110152","110413",..: 188 50 50 50 50 125 93 16 77 826 ...
##  $ fare     : num  211 152 152 152 152 ...
##  $ cabin    : Factor w/ 187 levels "","A10","A11",..: 45 81 81 81 81 151 147 17 63 1 ...
##  $ embarked : Factor w/ 4 levels "","C","Q","S": 4 4 4 4 4 4 4 4 4 2 ...
##  $ boat     : Factor w/ 28 levels "","1","10","11",..: 13 4 1 1 1 14 3 1 28 1 ...
##  $ body     : int  NA NA NA 135 NA NA NA NA NA 22 ...
##  $ home.dest: Factor w/ 370 levels "","?Havana, Cuba",..: 310 232 232 232 232 238 163 25 23 230 ...
```
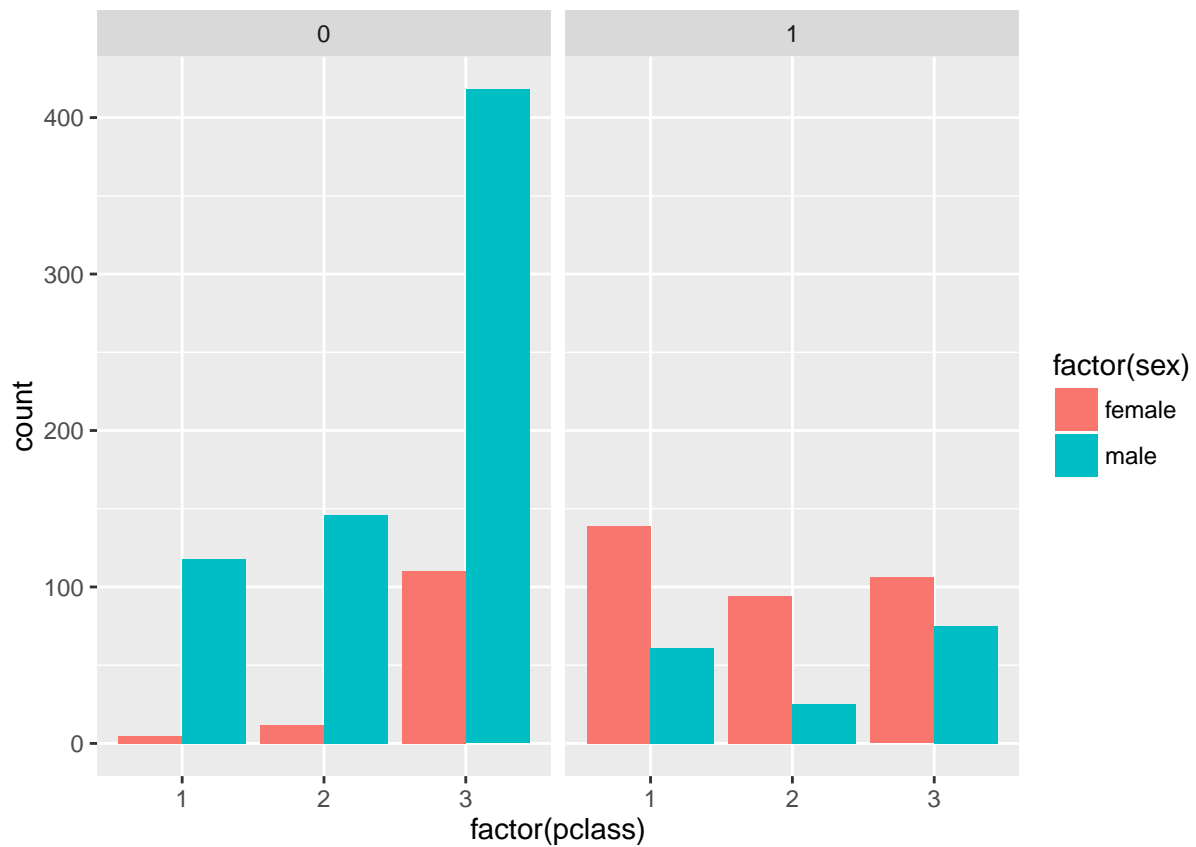
**ggplot() for the first instruction**

```
ggplot(titanic, aes(x=factor(pclass), fill=factor(sex))) + geom_bar(position="dodge")
```
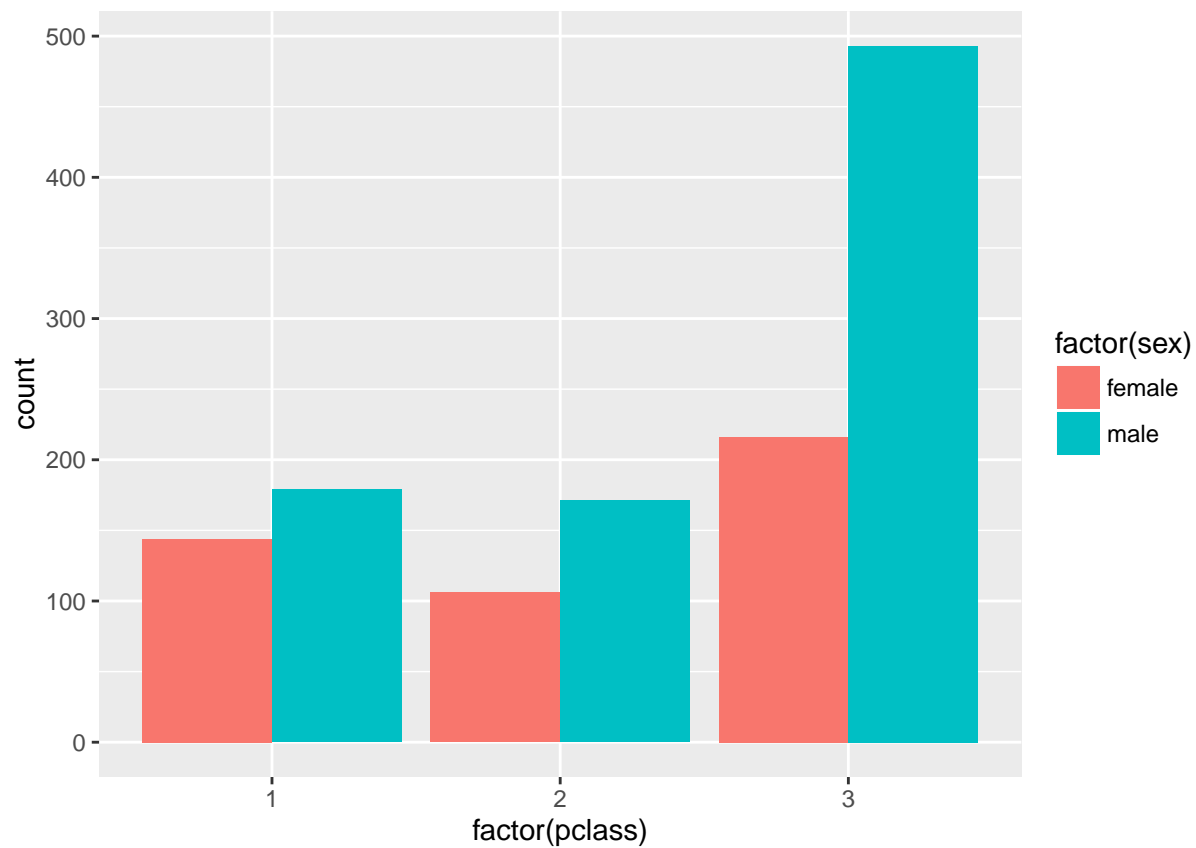


**ggplot() for the second instruction**

```
ggplot(titanic, aes(x=factor(pclass), fill=factor(sex))) + geom_bar(position="dodge") +
  facet_grid(".~survived")
```

**Position jitter (use below)**

```
posn.j <- position_jitter(0.5, 0)
```

**ggplot() for the last instruction**

```
ggplot(titanic, aes(x=factor(pclass), y=age, col=factor(sex))) +
  geom_jitter(position=posn.j, size=3, alpha=0.5) + facet_grid(".~survived")
```

```
## Warning: Removed 263 rows containing missing values (geom_point).
```