DV_Project1

Panyu Peng & Bryan Ho

September 12, 2015

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```r
require("jsonlite")
```

```
## Loading required package: jsonlite
##
## Attaching package: 'jsonlite'
##
## The following object is masked from 'package:utils':
##
##     View
```
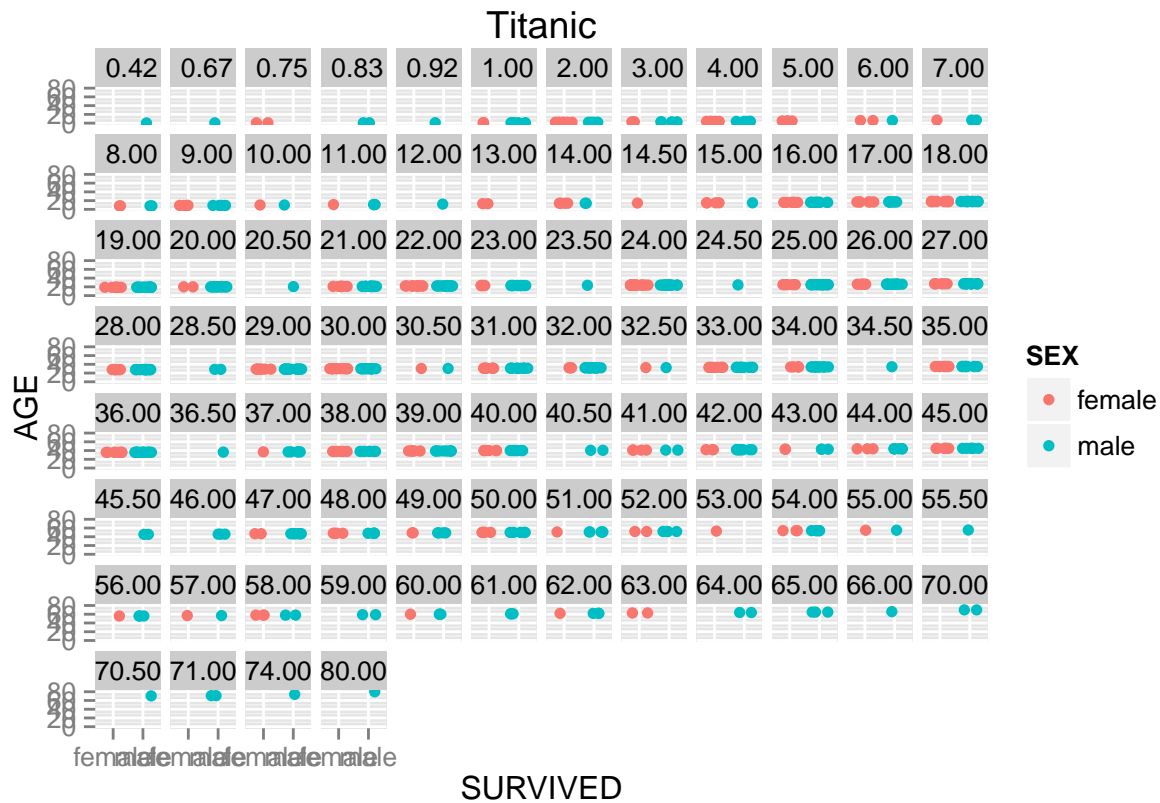
```r
require("RCurl")
```

```
## Loading required package: RCurl
## Loading required package: bitops
```

```r
require("ggplot2")
```

```
## Loading required package: ggplot2
```

```r
# Change the USER and PASS below to be your UTEid
df <- data.frame(fromJSON(getURL(URLencode('129.152.144.84:5001/rest/native/?query="select * from titan
ggplot()+
  coord_cartesian() +
  scale_x_discrete() +
  scale_y_continuous() +
  facet_wrap(~AGE, ncol=12)+
  labs(title= "Titanic")+
  labs(x="SURVIVED", y="AGE")+
  layer(data=df, mapping = aes(x=SEX, y=as.numeric(as.character(AGE)), color=SEX),
        stat="identity",
        stat_params=list(),
        geom="point",
        geom_params=list(),
        position=position_jitter(width=0.3, height=0)

        )
```

Titanic

AGE

SEX
● female
● male

0.42	0.67	0.75	0.83	0.92	1.00	2.00	3.00	4.00	5.00	6.00	7.00
8.00	9.00	10.00	11.00	12.00	13.00	14.00	14.50	15.00	16.00	17.00	18.00
19.00	20.00	20.50	21.00	22.00	23.00	23.50	24.00	24.50	25.00	26.00	27.00
28.00	28.50	29.00	30.00	30.50	31.00	32.00	32.50	33.00	34.00	34.50	35.00
36.00	36.50	37.00	38.00	39.00	40.00	40.50	41.00	42.00	43.00	44.00	45.00
45.50	46.00	47.00	48.00	49.00	50.00	51.00	52.00	53.00	54.00	55.00	55.50
56.00	57.00	58.00	59.00	60.00	61.00	62.00	63.00	64.00	65.00	66.00	70.00
70.50	71.00	74.00	80.00

SURVIVED

```r
# Change the USER and PASS below to be your UTEid
df <- data.frame(fromJSON(getURL(URLencode('129.152.144.84:5001/rest/native/?query="select * from titani

summary(df)
```

```
##   PASSENGERID      SURVIVED         PCLASS
## Min.   :  1.0   Min.   :0.0000   Min.   :1.000
## 1st Qu.:223.5   1st Qu.:0.0000   1st Qu.:2.000
## Median :446.0   Median :0.0000   Median :3.000
## Mean   :446.0   Mean   :0.3838   Mean   :2.309
## 3rd Qu.:668.5   3rd Qu.:1.0000   3rd Qu.:3.000
## Max.   :891.0   Max.   :1.0000   Max.   :3.000
##
##                                      NAME          SEX          AGE
## Abbing, Mr. Anthony                  :  1   female:314   null   :177
## Abbott, Mr. Rossmore Edward          :  1   male  :577   24     : 30
## Abbott, Mrs. Stanton (Rosa Hunt)     :  1                22     : 27
## Abelson, Mr. Samuel                  :  1                18     : 26
## Abelson, Mrs. Samuel (Hannah Wizosky):  1                19     : 25
## Adahl, Mr. Mauritz Nils Martin       :  1                28     : 25
## (Other)                              :885                (Other):581
##      SIBSP           PARCH            TICKET         FARE
## Min.   :0.000   Min.   :0.0000   1601   :  7   Min.   :  0.00
## 1st Qu.:0.000   1st Qu.:0.0000   347082 :  7   1st Qu.:  7.91
## Median :0.000   Median :0.0000   CA. 2343:  7   Median : 14.45
## Mean   :0.523   Mean   :0.3816   3101295 :  6   Mean   : 32.20
## 3rd Qu.:1.000   3rd Qu.:0.0000   347088  :  6   3rd Qu.: 31.00
## Max.   :8.000   Max.   :6.0000   CA 2144 :  6   Max.   :512.33
```

```
##                                    (Other) :852
##           CABIN      EMBARKED
##   null        :687   C   :168
##   B96 B98     :  4   null:  2
##   C23 C25 C27:  4   Q   : 77
##   G6          :  4   S   :644
##   C22 C26     :  3
##   D           :  3
##   (Other)     :186
```

```r
head(df)
```

```
##    PASSENGERID SURVIVED PCLASS
## 1          414        0      2
## 2          415        1      3
## 3          416        0      3
## 4          417        1      2
## 5          418        1      2
## 6          419        0      2
##                                            NAME    SEX  AGE SIBSP PARCH
## 1              Cunningham, Mr. Alfred Fleming   male null     0     0
## 2                    Sundman, Mr. Johan Julian   male   44     0     0
## 3       Meek, Mrs. Thomas (Annie Louise Rowley) female null     0     0
## 4 Drew, Mrs. James Vivian (Lulu Thorne Christian) female   34     1     1
## 5             Silven, Miss. Lyyli Karoliina female   18     0     2
## 6                  Matthews, Mr. William John   male   30     0     0
##             TICKET    FARE CABIN EMBARKED
## 1           239853  0.000  null        S
## 2 STON/O 2. 3101269  7.925  null        S
## 3           343095  8.050  null        S
## 4            28220 32.500  null        S
## 5           250652 13.000  null        S
## 6            28228 13.000  null        S
```

```r
# Change the USER and PASS below to be your UTEid
df <- data.frame(fromJSON(getURL(URLencode('129.152.144.84:5001/rest/native/?query="select * from titani
#df
#summary(df)
#head(df)

require(extrafont)
```

```
## Loading required package: extrafont
## Registering fonts with R
```

```r
ggplot() +
  coord_cartesian() +
  scale_x_continuous() +
  scale_y_continuous() +
  labs(title='Titanic') +
  labs(x="Age", y=paste("Fare")) +
  layer(data=df,
```

```
        mapping=aes(x=as.numeric(as.character(AGE)), y=as.numeric(as.character(FARE)), color=SEX),
        stat="identity",
        stat_params=list(),
        geom="point",
        geom_params=list(),
        position=position_jitter(width=0.3, height=0)
    )
```
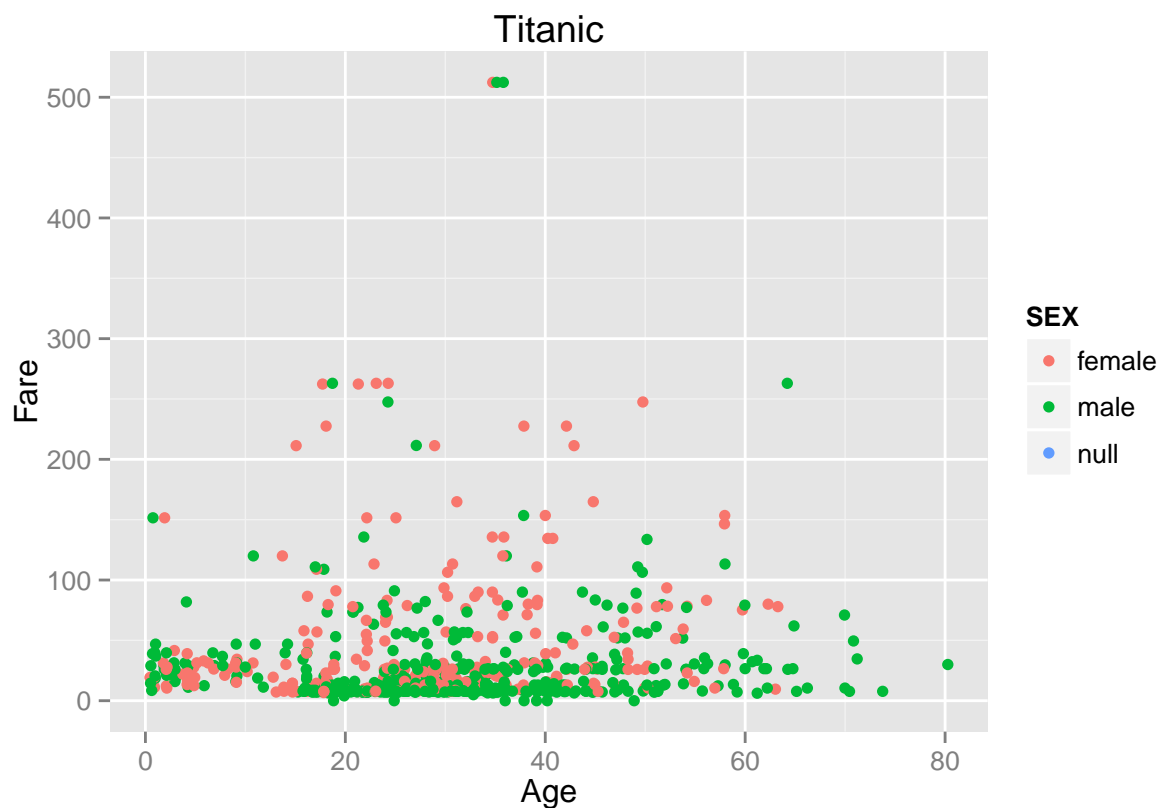
```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: Removed 179 rows containing missing values (geom_point).
```



```
require("jsonlite")
require("RCurl")
require("ggplot2")
# Change the USER and PASS below to be your UTEid
df <- data.frame(fromJSON(getURL(URLencode('129.152.144.84:5001/rest/native/?query="select * from titan
#df
#summary(df)
#head(df)

require(extrafont)
ggplot() +
  coord_cartesian() +
```
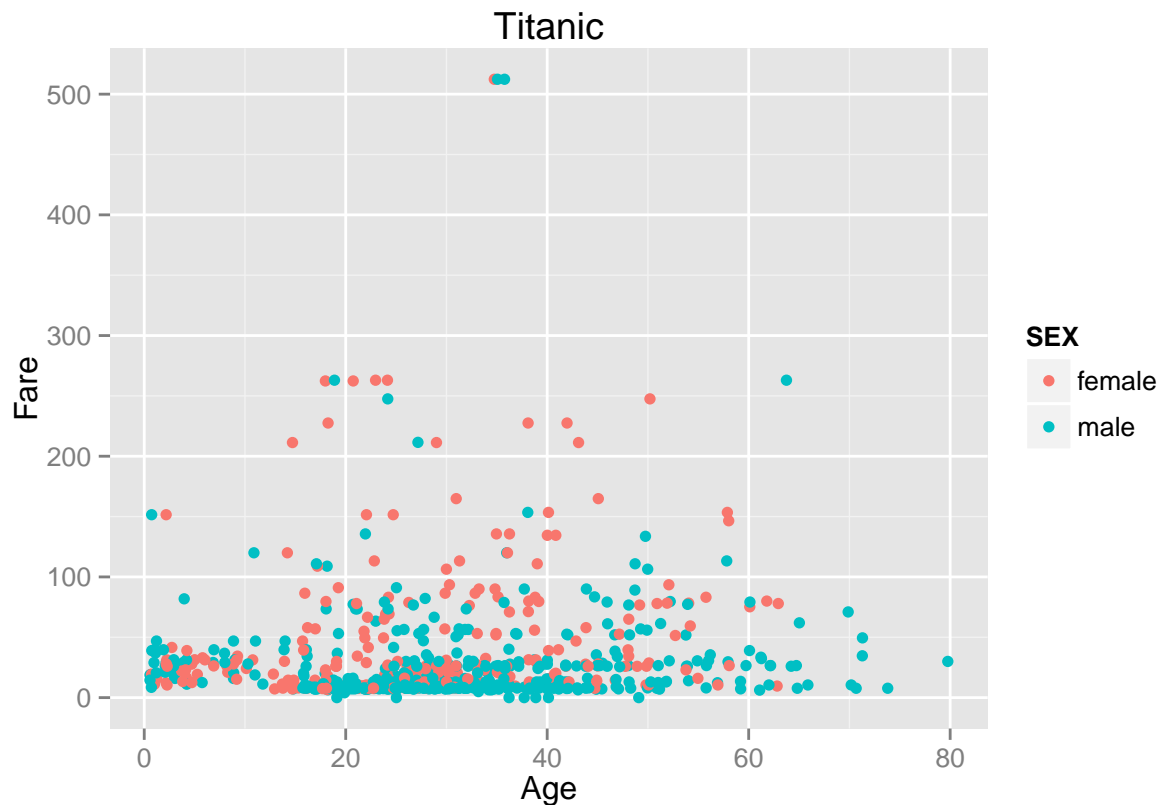
```
  scale_x_continuous() +
  scale_y_continuous() +
  #facet_wrap(~SURVIVED) +
  #facet_grid(.~SURVIVED, labeller=label_both) + # Same as facet_wrap but with a label.
  #facet_grid(PCLASS~SURVIVED, labeller=label_both) +
  labs(title='Titanic') +
  labs(x="Age", y=paste("Fare")) +
  layer(data=df,
        mapping=aes(x=as.numeric(as.character(AGE)), y=as.numeric(as.character(FARE)), color=SEX),
        stat="identity",
        stat_params=list(),
        geom="point",
        geom_params=list(),
        #position=position_identity()
        position=position_jitter(width=0.3, height=0)
  )
```

Warning: NAs introduced by coercion

Warning: Removed 177 rows containing missing values (geom_point).
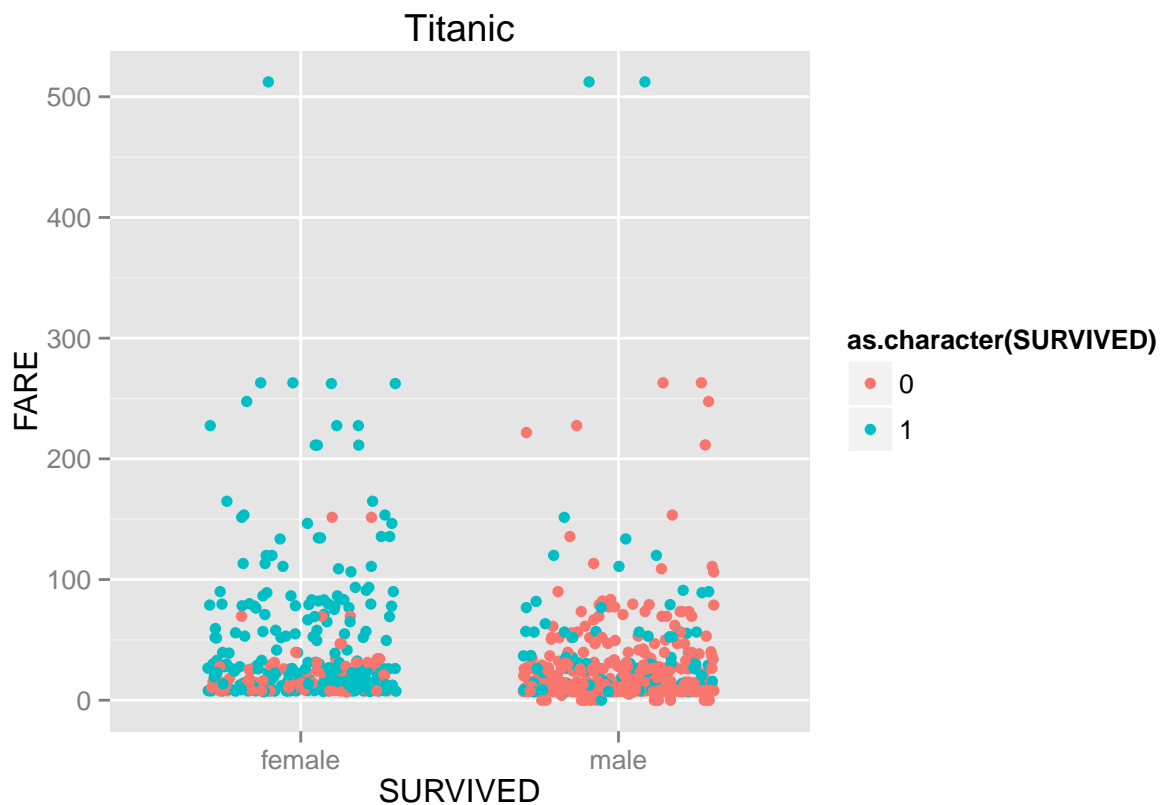


```
ggplot() +
  coord_cartesian() +
  scale_x_discrete() +
  scale_y_continuous() +
  #facet_grid(PCLASS~SURVIVED, labeller=label_both) +
```

```
    labs(title='Titanic') +
    labs(x="SURVIVED", y=paste("FARE")) +
    layer(data=df,
          mapping=aes(x=SEX, y=as.numeric(as.character(FARE)), color=as.character(SURVIVED)),
          stat="identity",
          stat_params=list(),
          geom="point",
          geom_params=list(),
          #position=position_identity()
          position=position_jitter(width=0.3, height=0)
    )
```
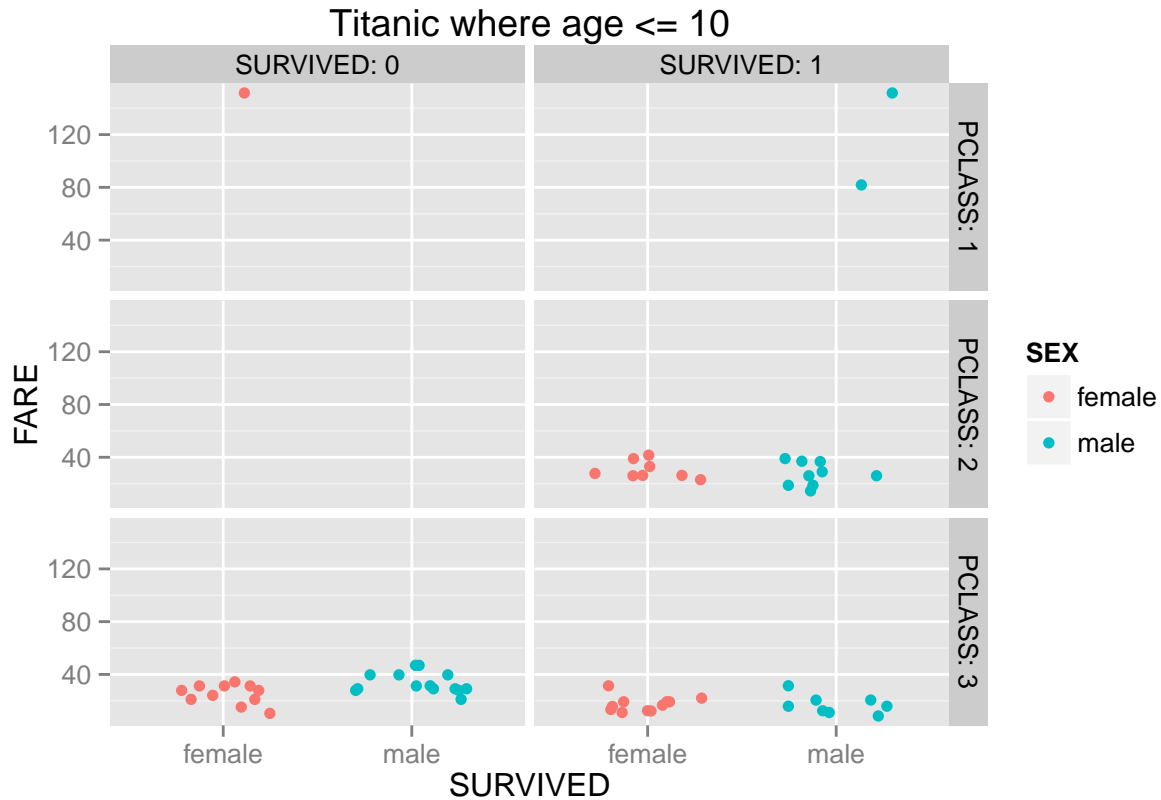


```
require("jsonlite")
require("RCurl")
require("ggplot2")
# Change the USER and PASS below to be your UTEid
df <- data.frame(fromJSON(getURL(URLencode('129.152.144.84:5001/rest/native/?query="select * from titan
#df
#summary(df)
#head(df)
ggplot() +
  coord_cartesian() +
  scale_x_discrete() +
  scale_y_continuous() +
  facet_grid(PCLASS~SURVIVED, labeller=label_both) +
  labs(title='Titanic where age <= 10') +
  labs(x="SURVIVED", y=paste("FARE")) +
```
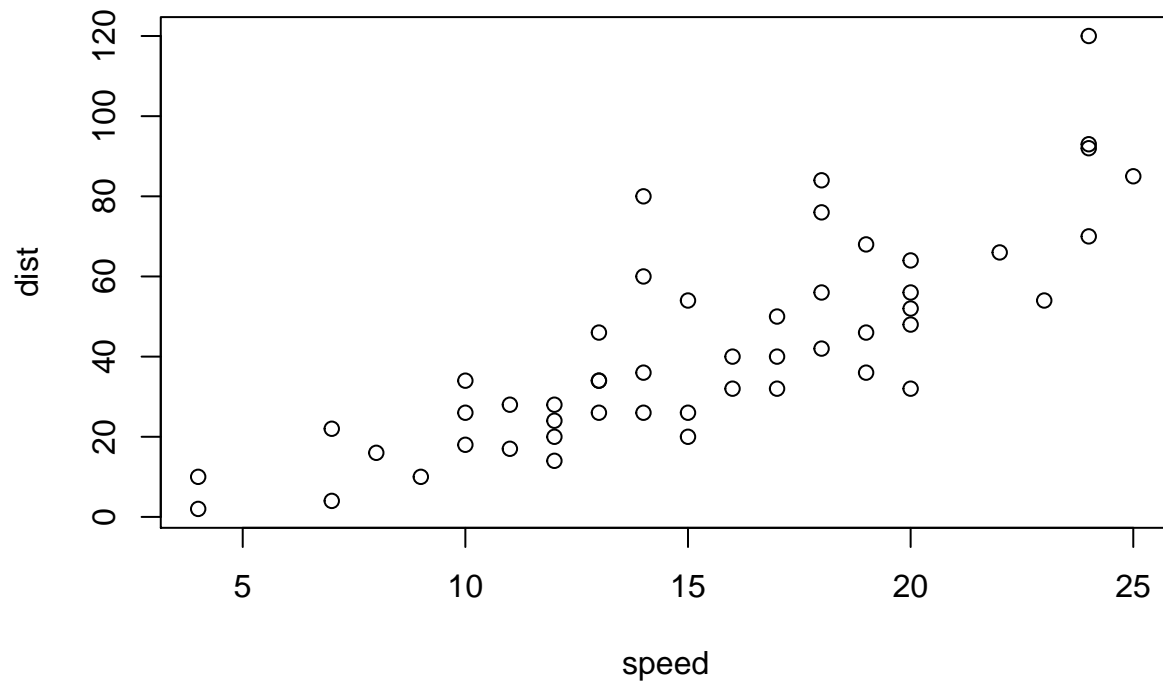
```
layer(data=df,
      mapping=aes(x=SEX, y=as.numeric(as.character(FARE)), color=SEX),
      stat="identity",
      stat_params=list(),
      geom="point",
      geom_params=list(),
      #position=position_identity()
      position=position_jitter(width=0.3, height=0)
)
```


Titanic where age <= 10

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.