ggplot2

Daniela Pinto Veizaga

13/7/2019

Reference Books

1) Cookbook for R: http://www.cookbook-r.com/

```
library(ggplot2)
```

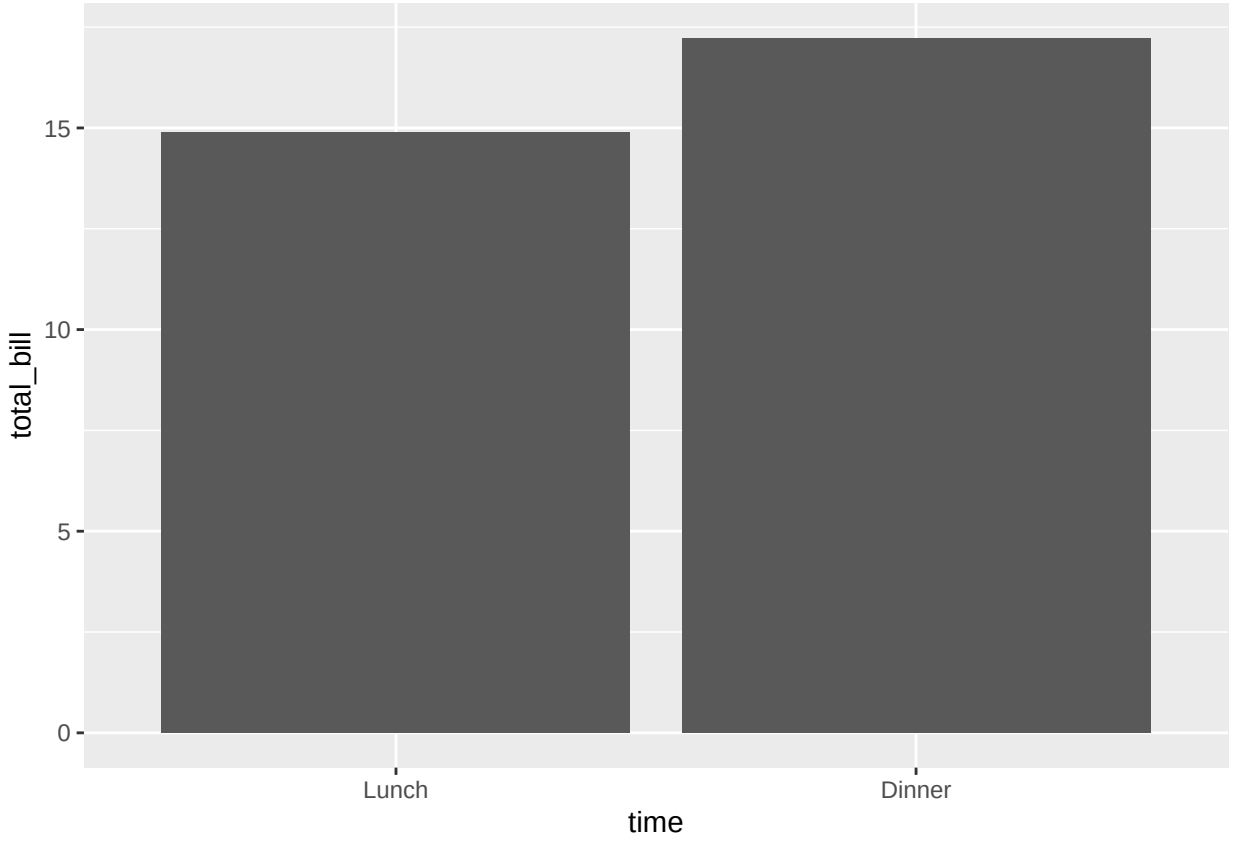
```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
dat <- data.frame(
  time = factor(c("Lunch","Dinner"), levels=c("Lunch","Dinner")),
  total_bill = c(14.89, 17.23)
)
dat
```

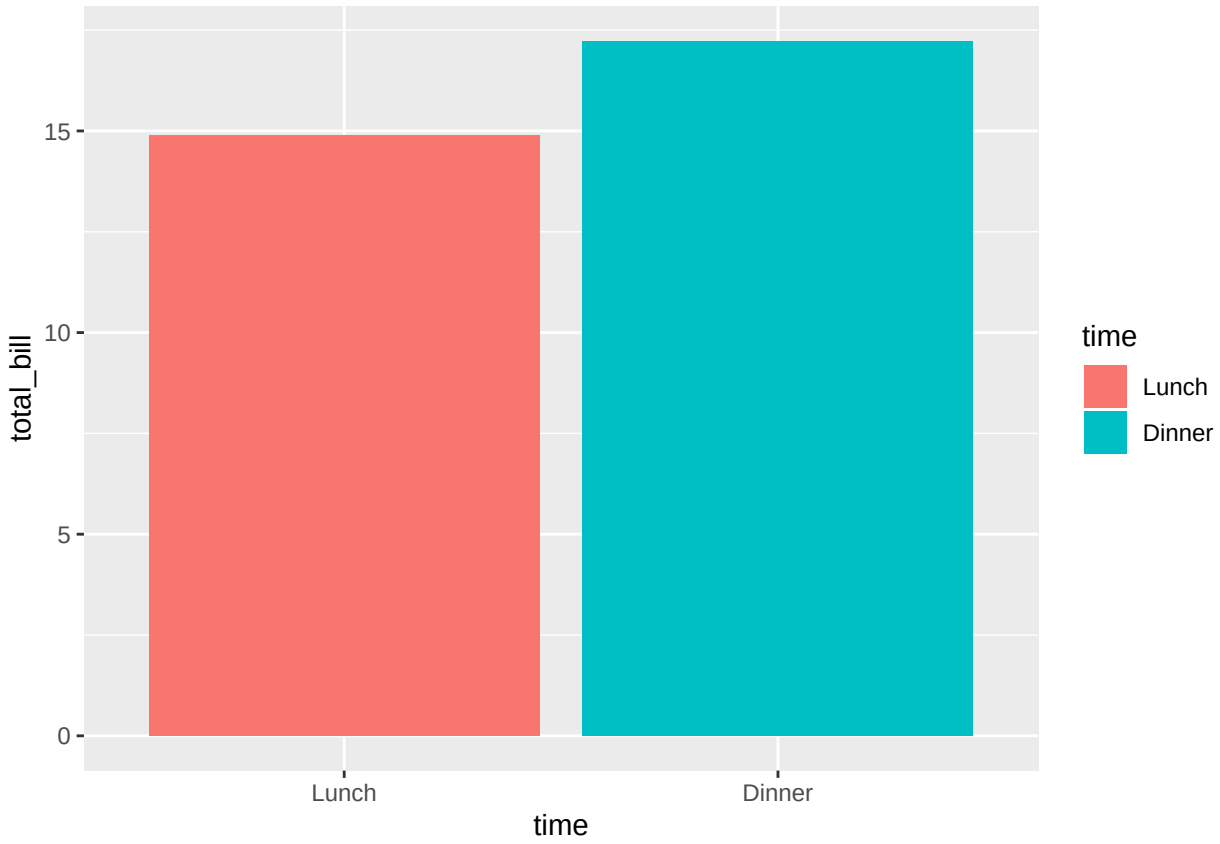
```
##     time total_bill
## 1  Lunch      14.89
## 2 Dinner      17.23
```

```
#>     time total_bill
#> 1  Lunch      14.89
#> 2 Dinner      17.23

# Load the ggplot2 package
```

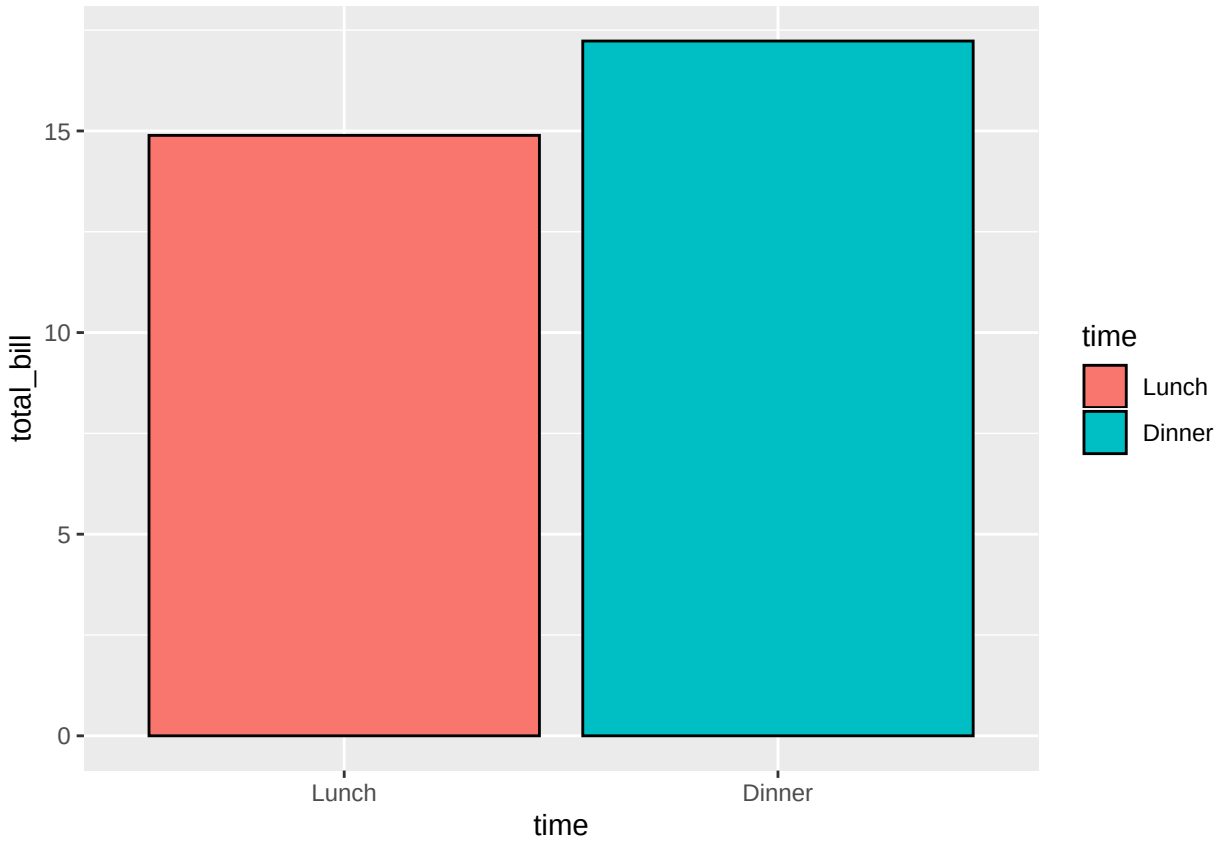
```
# Very basic bar graph
ggplot(data=dat, aes(x=time, y=total_bill)) +
    geom_bar(stat="identity")
```



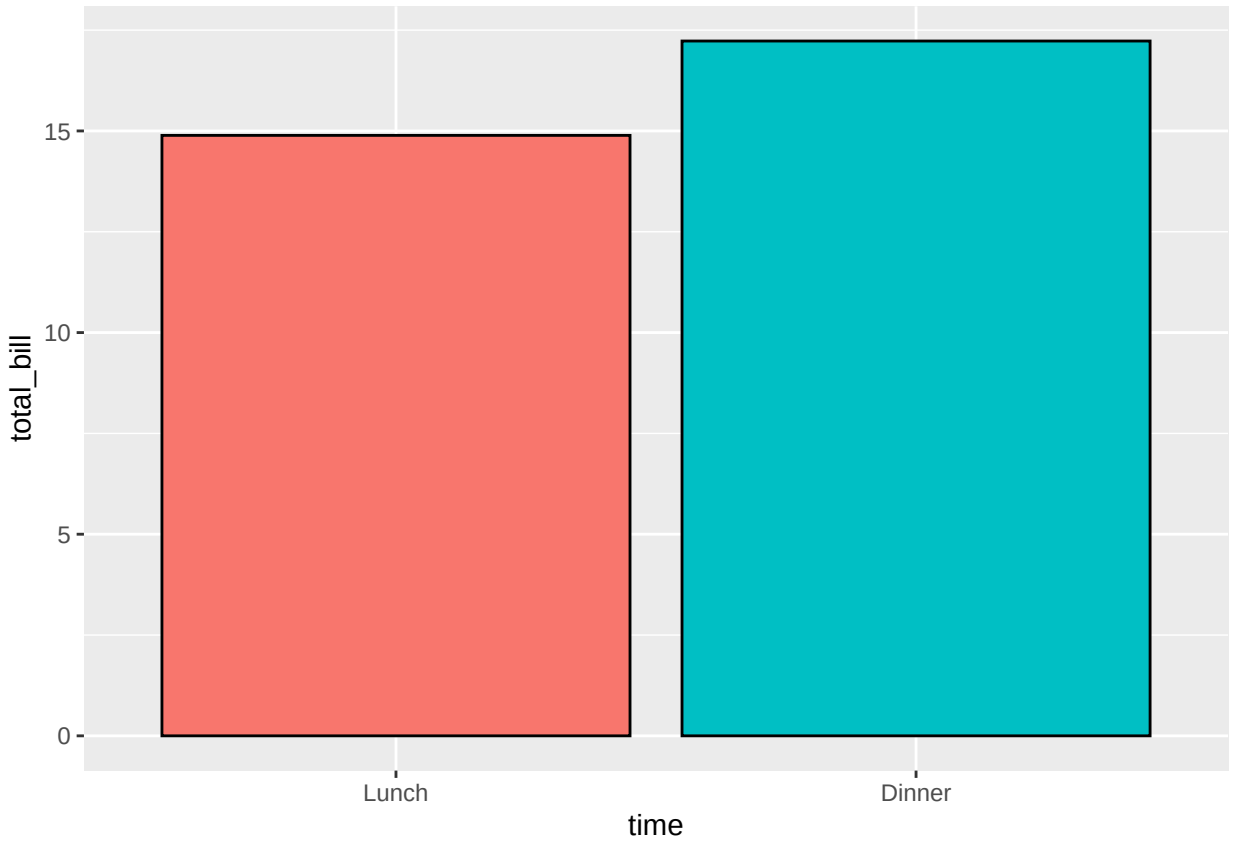

```
# Map the time of day to different fill colors
ggplot(data=dat, aes(x=time, y=total_bill, fill=time)) +
    geom_bar(stat="identity")
```



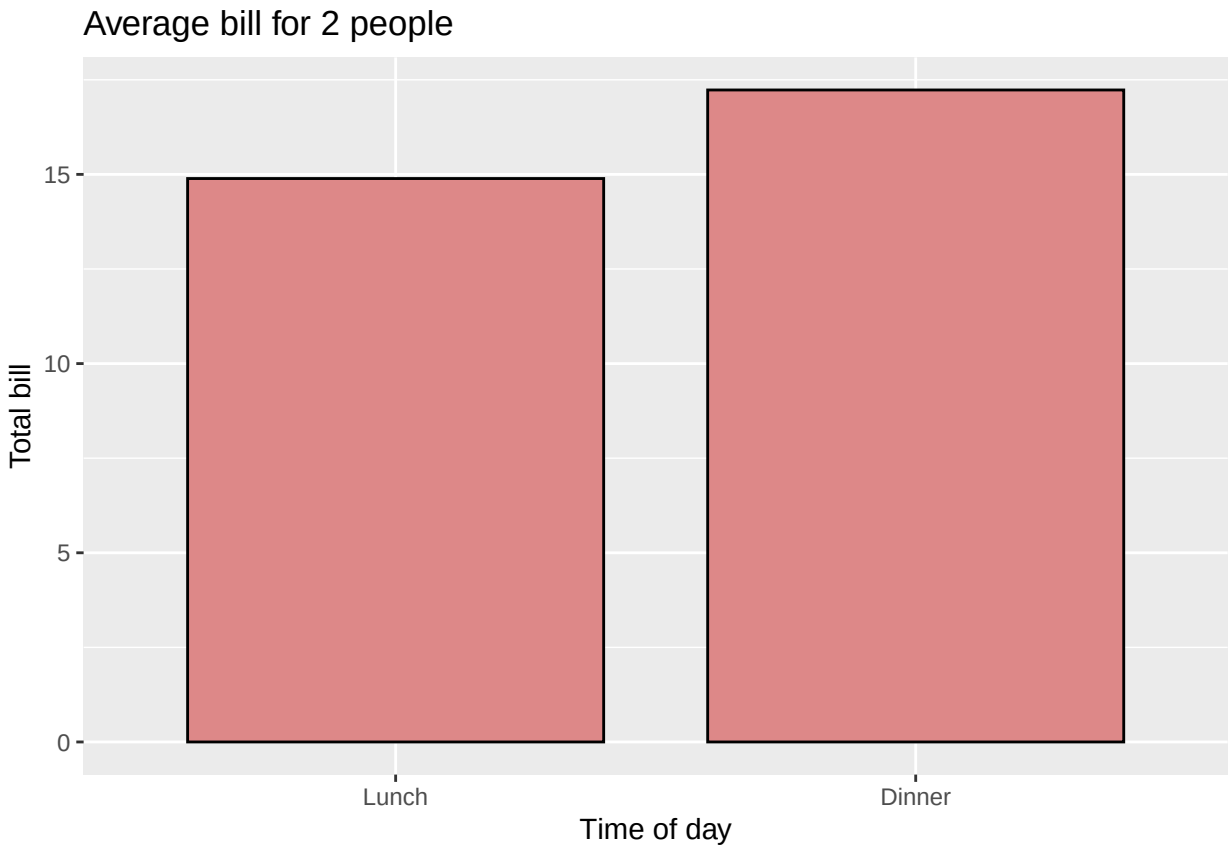

```
## This would have the same result as above
# ggplot(data=dat, aes(x=time, y=total_bill)) +
#     geom_bar(aes(fill=time), stat="identity")


# Add a black outline
ggplot(data=dat, aes(x=time, y=total_bill, fill=time)) +
    geom_bar(colour="black", stat="identity")
```




```
# No legend, since the information is redundant
ggplot(data=dat, aes(x=time, y=total_bill, fill=time)) +
    geom_bar(colour="black", stat="identity") +
    guides(fill=FALSE)
```




```
# Add title, narrower bars, fill color, and change axis labels
ggplot(data=dat, aes(x=time, y=total_bill, fill=time)) +
    geom_bar(colour="black", fill="#DD8888", width=.8, stat="identity") +
    guides(fill=FALSE) +
    xlab("Time of day") + ylab("Total bill") +
    ggtitle("Average bill for 2 people")
```

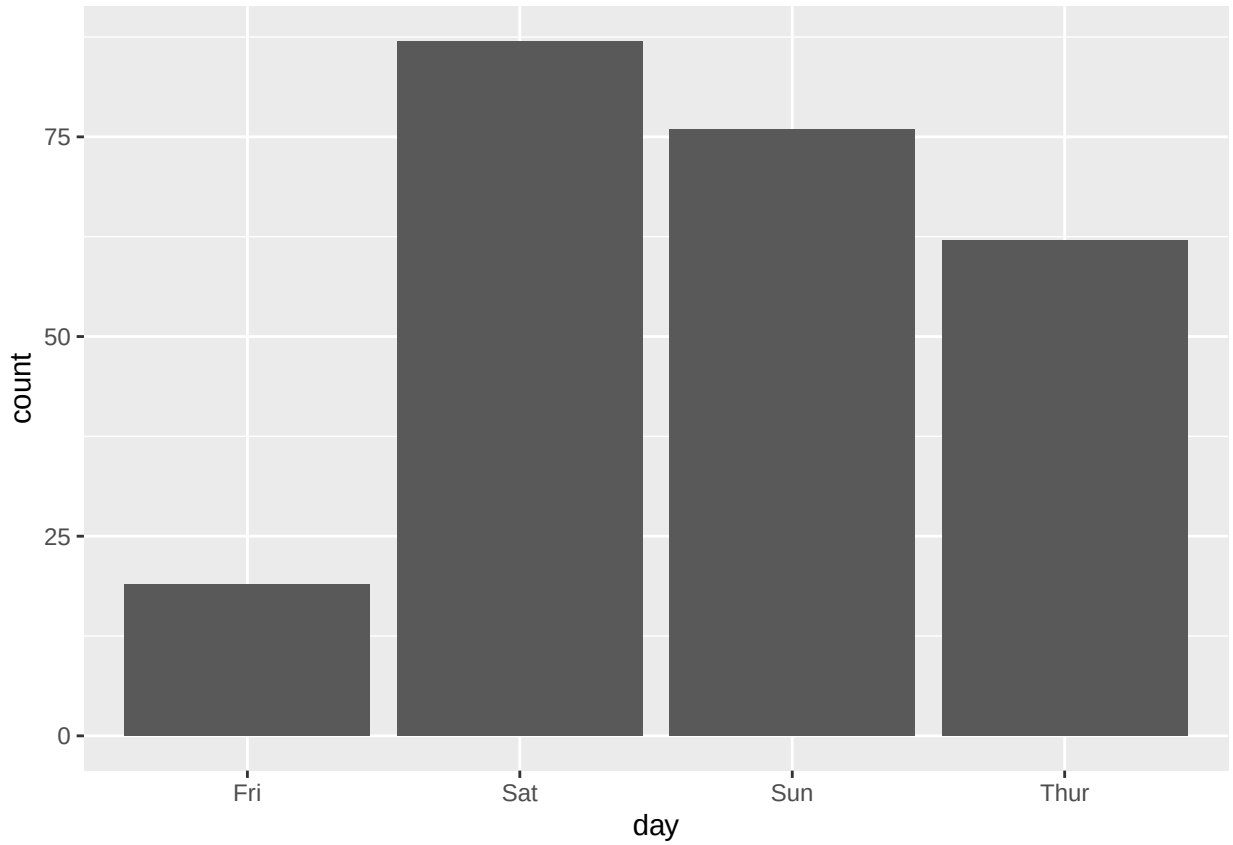



```
library(reshape2)
# Look at fist several rows
head(tips)
```

```
##   total_bill  tip    sex smoker day   time size
## 1      16.99 1.01 Female     No Sun Dinner    2
## 2      10.34 1.66   Male     No Sun Dinner    3
## 3      21.01 3.50   Male     No Sun Dinner    3
## 4      23.68 3.31   Male     No Sun Dinner    2
## 5      24.59 3.61 Female     No Sun Dinner    4
## 6      25.29 4.71   Male     No Sun Dinner    4
```

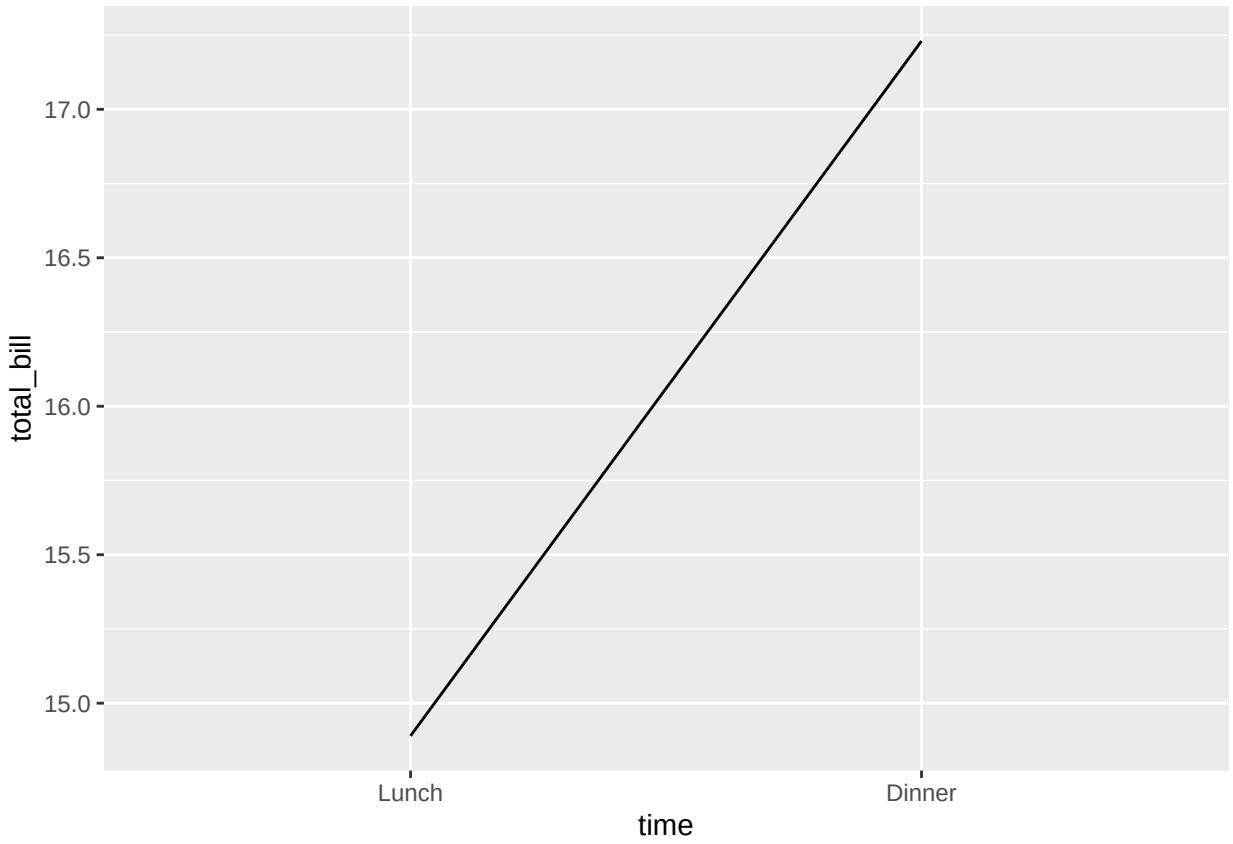
```
#>   total_bill  tip    sex smoker day   time size
#> 1      16.99 1.01 Female     No Sun Dinner    2
#> 2      10.34 1.66   Male     No Sun Dinner    3
#> 3      21.01 3.50   Male     No Sun Dinner    3
#> 4      23.68 3.31   Male     No Sun Dinner    2
#> 5      24.59 3.61 Female     No Sun Dinner    4
#> 6      25.29 4.71   Male     No Sun Dinner    4
```

```
# Bar graph of counts
ggplot(data=tips, aes(x=day)) +
    geom_bar(stat="count")
```

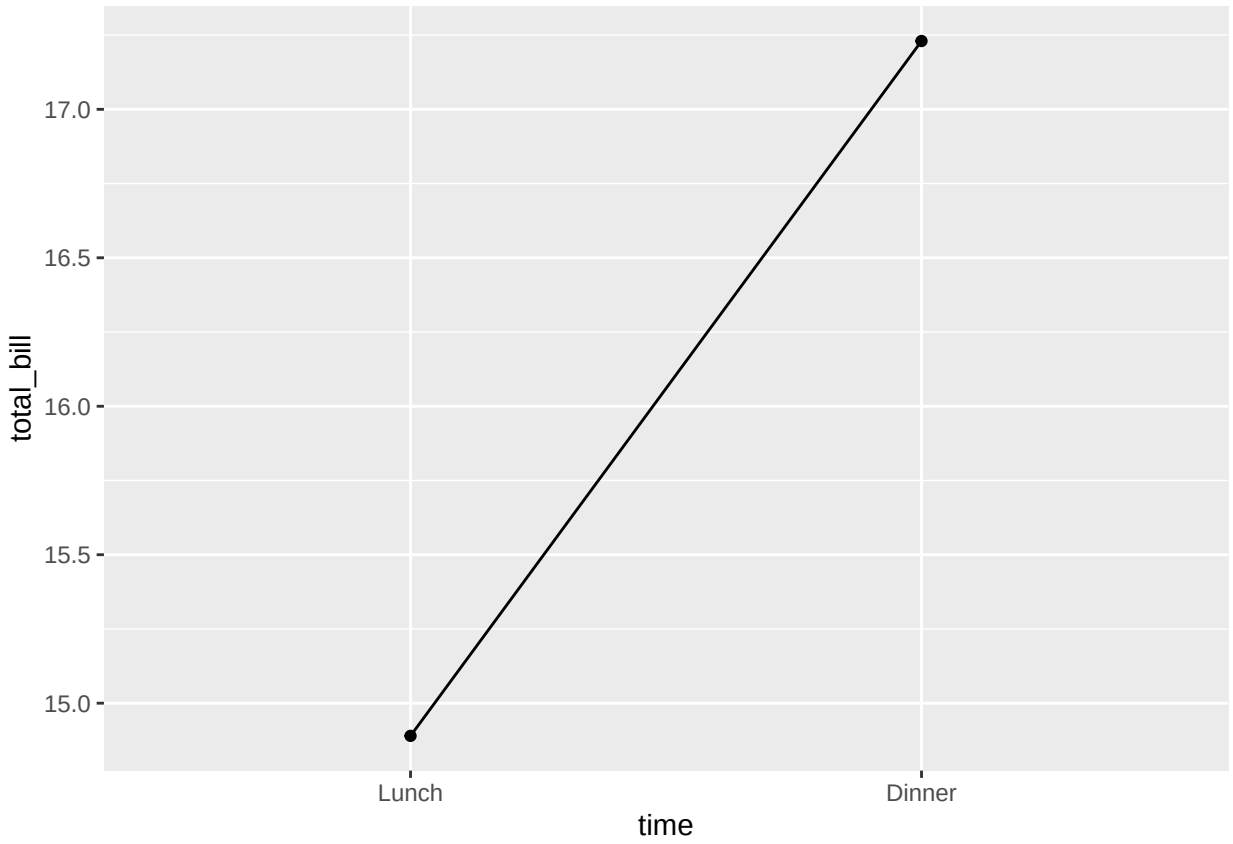



```
## Equivalent to this, since stat="bin" is the default:
# ggplot(data=tips, aes(x=day)) +
#    geom_bar()
```

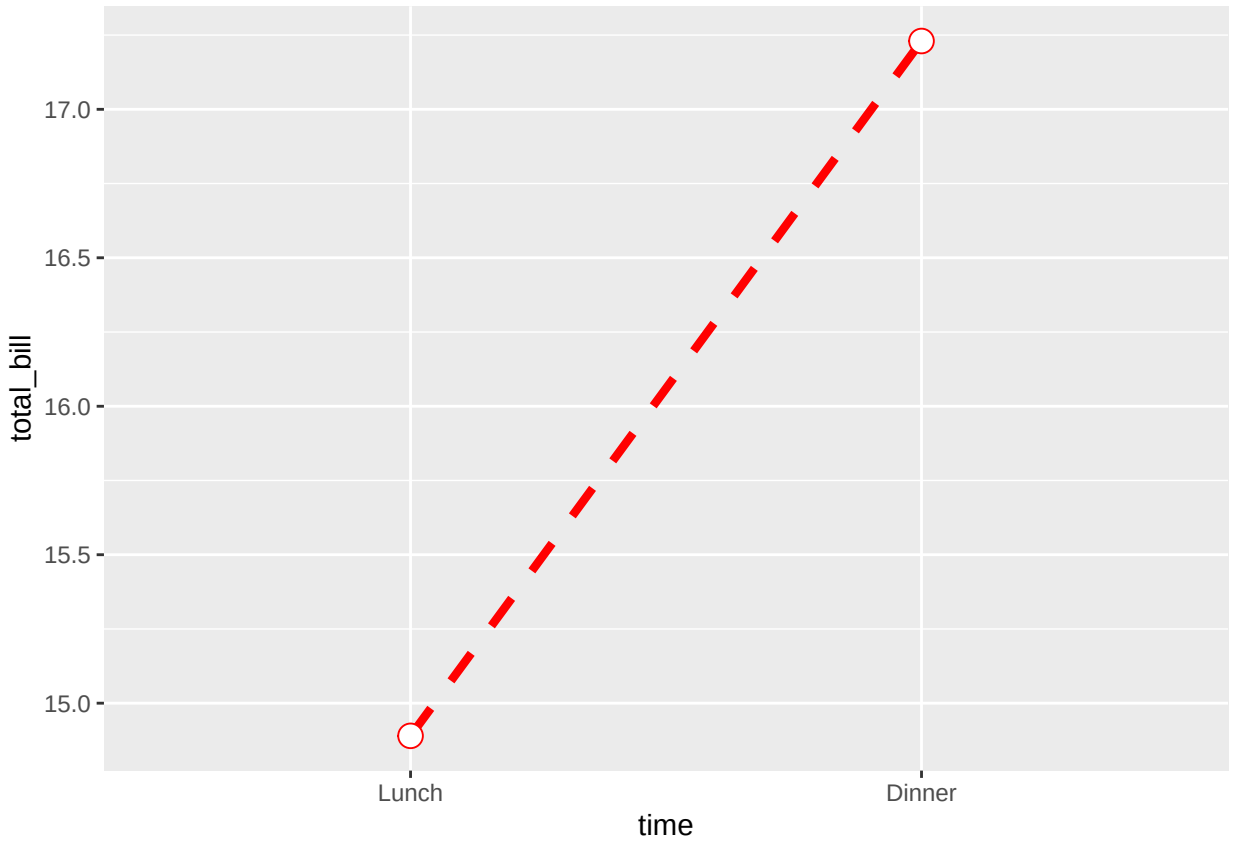
```
# Basic line graph
ggplot(data=dat, aes(x=time, y=total_bill, group=1)) +
    geom_line()
```



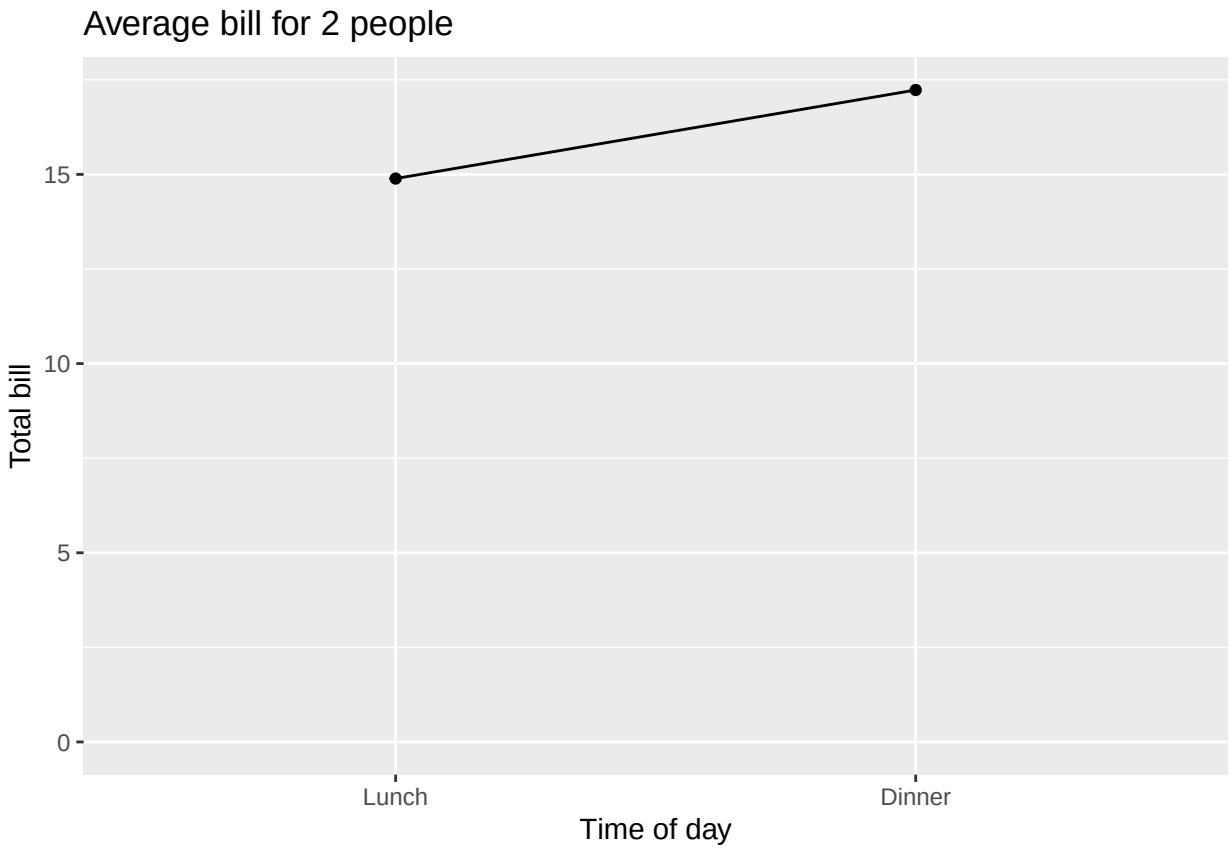

```
## This would have the same result as above
# ggplot(data=dat, aes(x=time, y=total_bill)) +
#     geom_line(aes(group=1))

# Add points
ggplot(data=dat, aes(x=time, y=total_bill, group=1)) +
    geom_line() +
    geom_point()
```




```
# Change color of both line and points
# Change line type and point type, and use thicker line and larger points
# Change points to circles with white fill
ggplot(data=dat, aes(x=time, y=total_bill, group=1)) +
    geom_line(colour="red", linetype="dashed", size=1.5) +
    geom_point(colour="red", size=4, shape=21, fill="white")
```




```
# Change the y-range to go from 0 to the maximum value in the total_bill column,
# and change axis labels
ggplot(data=dat, aes(x=time, y=total_bill, group=1)) +
    geom_line() +
    geom_point() +
    expand_limits(y=0) +
    xlab("Time of day") + ylab("Total bill") +
    ggtitle("Average bill for 2 people")
```

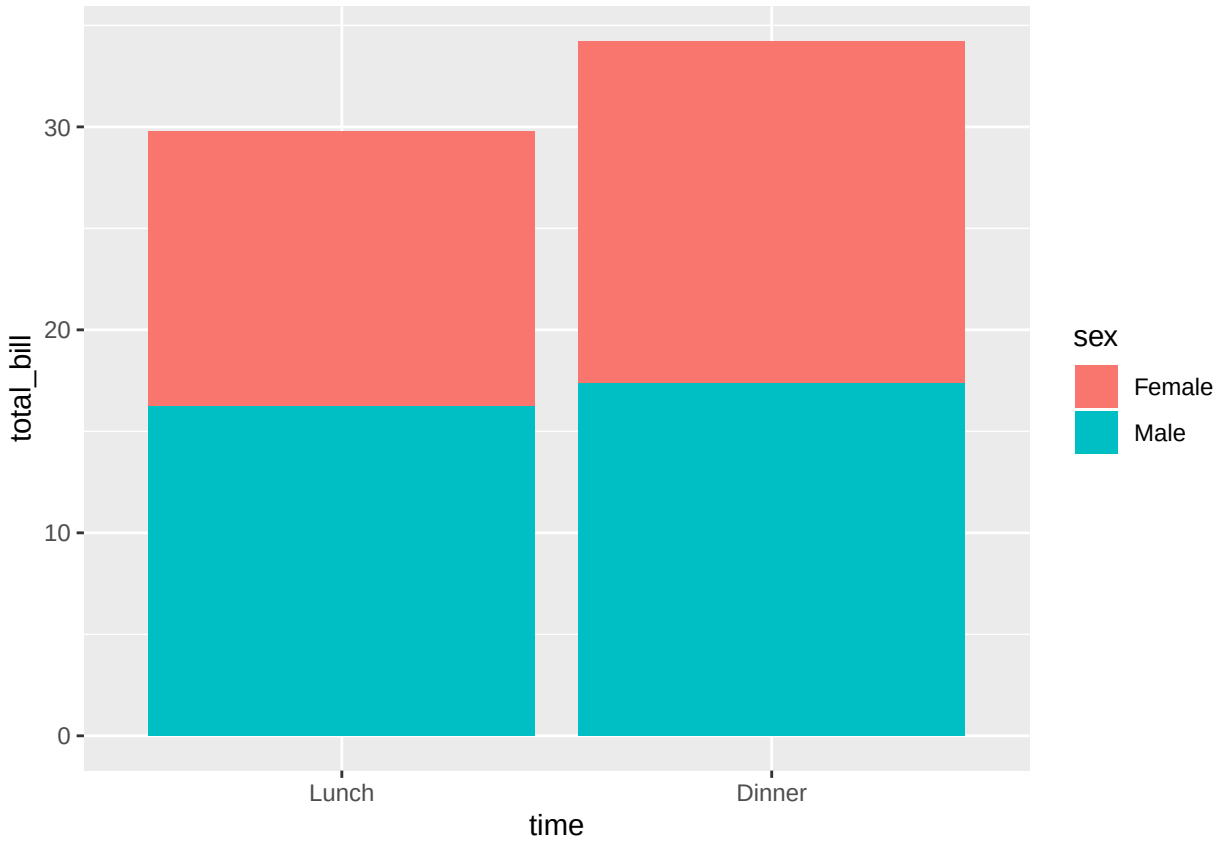



```
dat1 <- data.frame(
    sex = factor(c("Female","Female","Male","Male")),
    time = factor(c("Lunch","Dinner","Lunch","Dinner"), levels=c("Lunch","Dinner")),
    total_bill = c(13.53, 16.81, 16.24, 17.42)
)
dat1
```

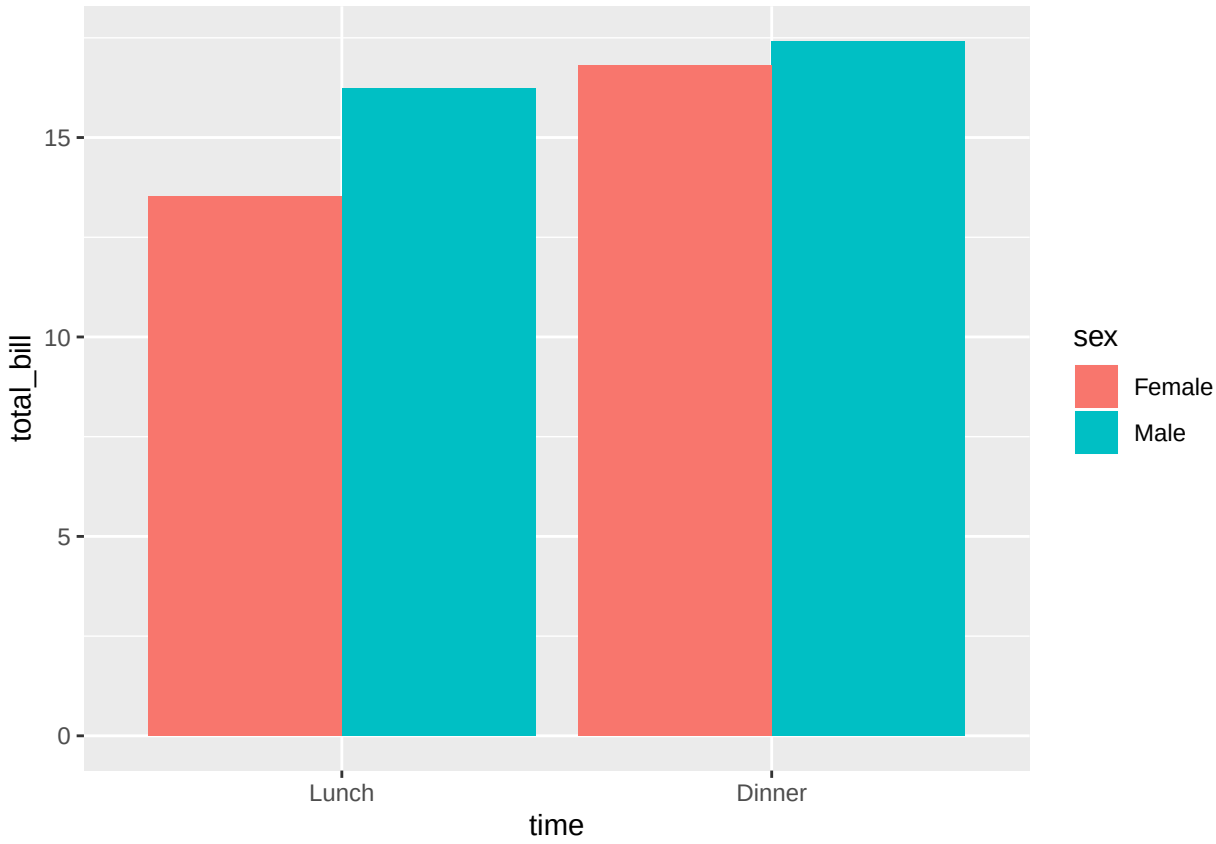
```
##      sex   time total_bill
## 1 Female  Lunch      13.53
## 2 Female Dinner      16.81
## 3   Male  Lunch      16.24
## 4   Male Dinner      17.42
```

```
#>      sex   time total_bill
#> 1 Female  Lunch      13.53
#> 2 Female Dinner      16.81
#> 3   Male  Lunch      16.24
#> 4   Male Dinner      17.42
```

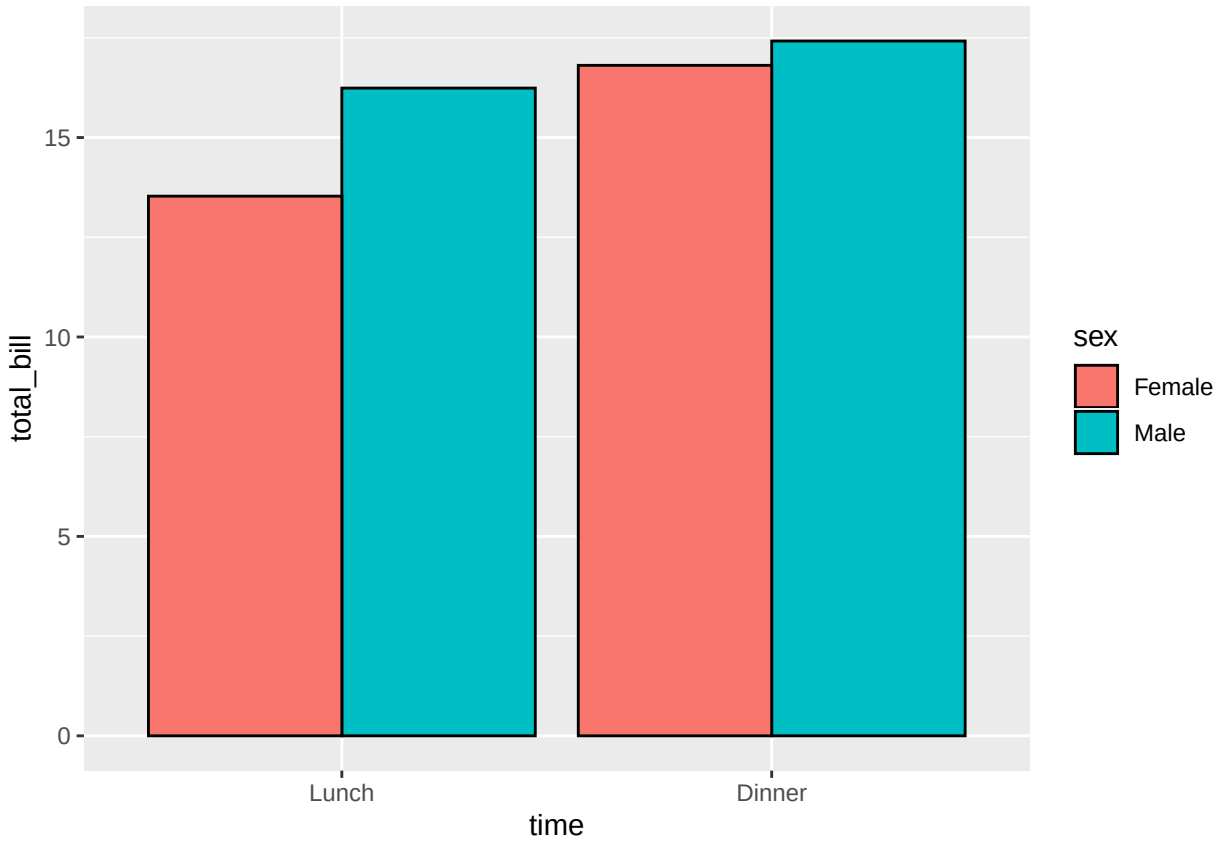
```
# Stacked bar graph -- this is probably not what you want
ggplot(data=dat1, aes(x=time, y=total_bill, fill=sex)) +
    geom_bar(stat="identity")
```



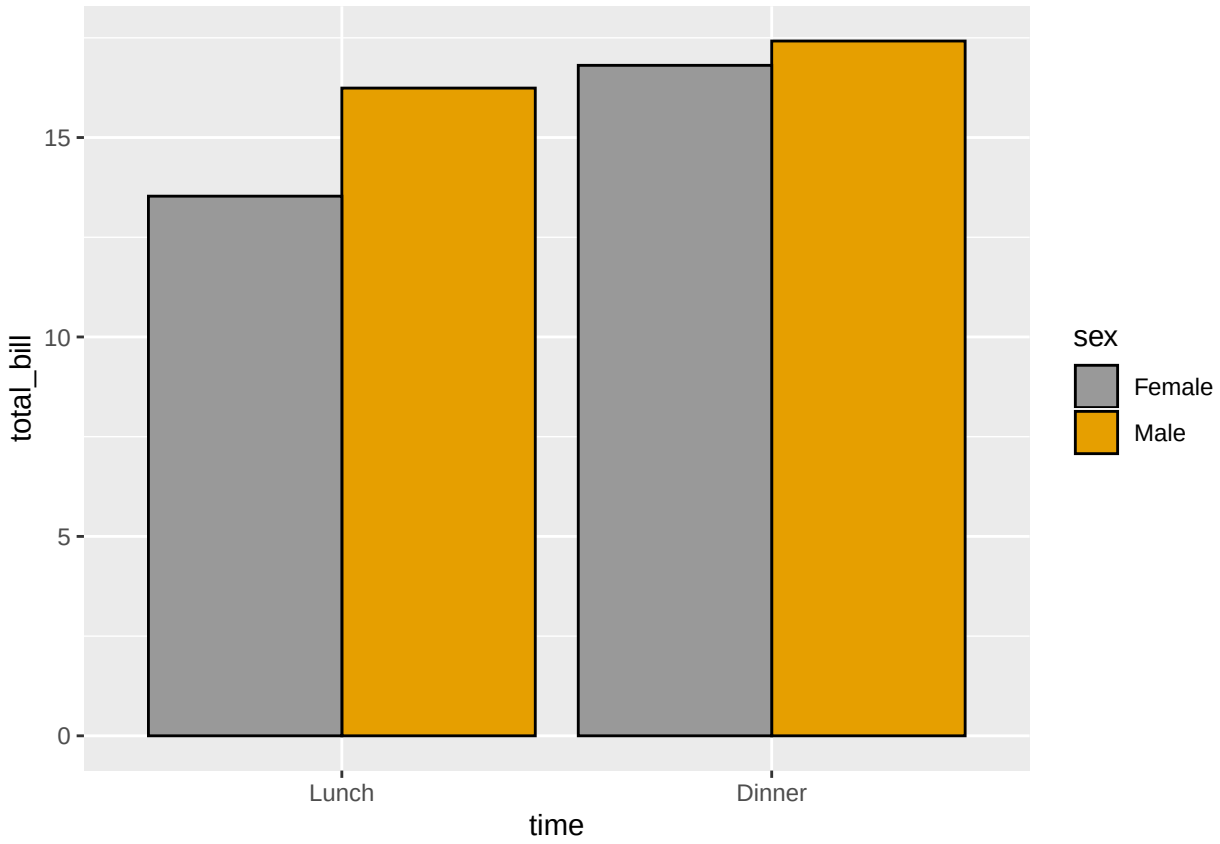

```
# Bar graph, time on x-axis, color fill grouped by sex -- use position_dodge()
ggplot(data=dat1, aes(x=time, y=total_bill, fill=sex)) +
    geom_bar(stat="identity", position=position_dodge())
```



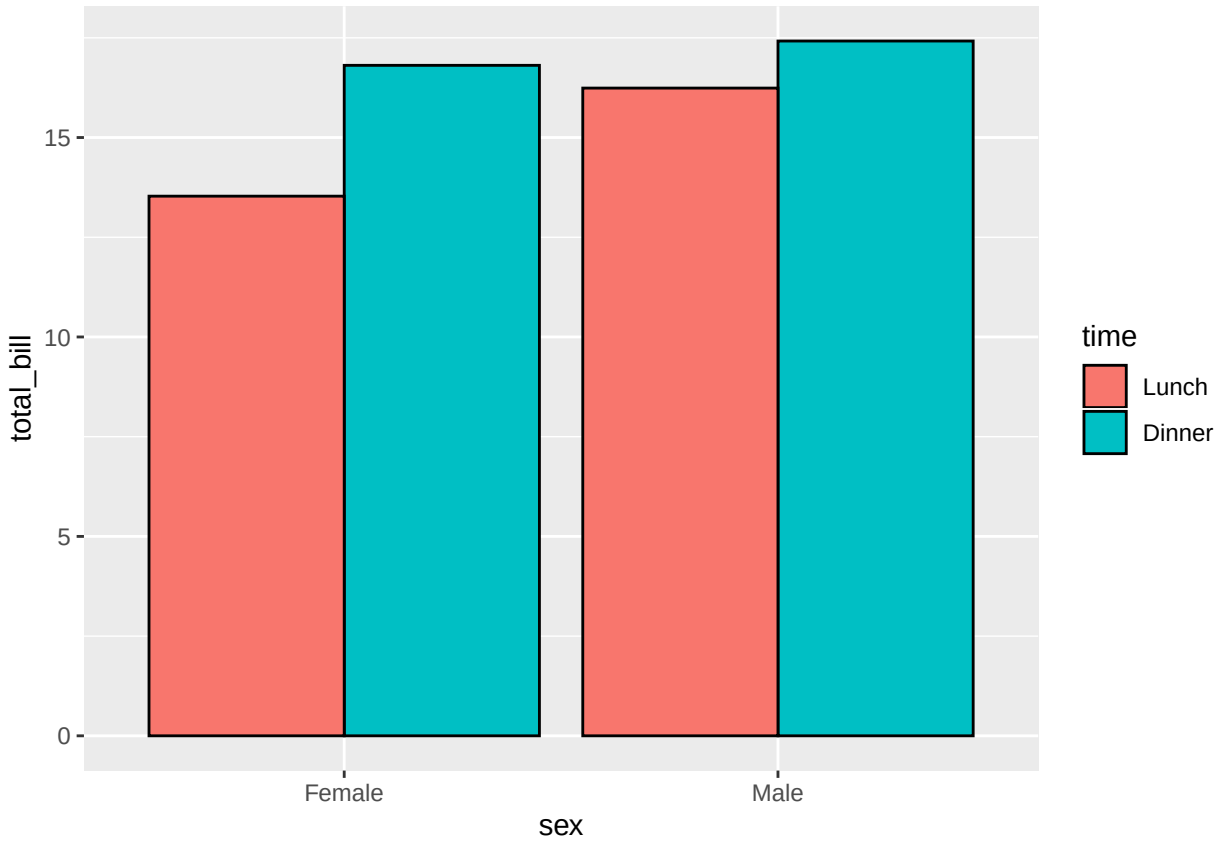

```
ggplot(data=dat1, aes(x=time, y=total_bill, fill=sex)) +
    geom_bar(stat="identity", position=position_dodge(), colour="black")
```



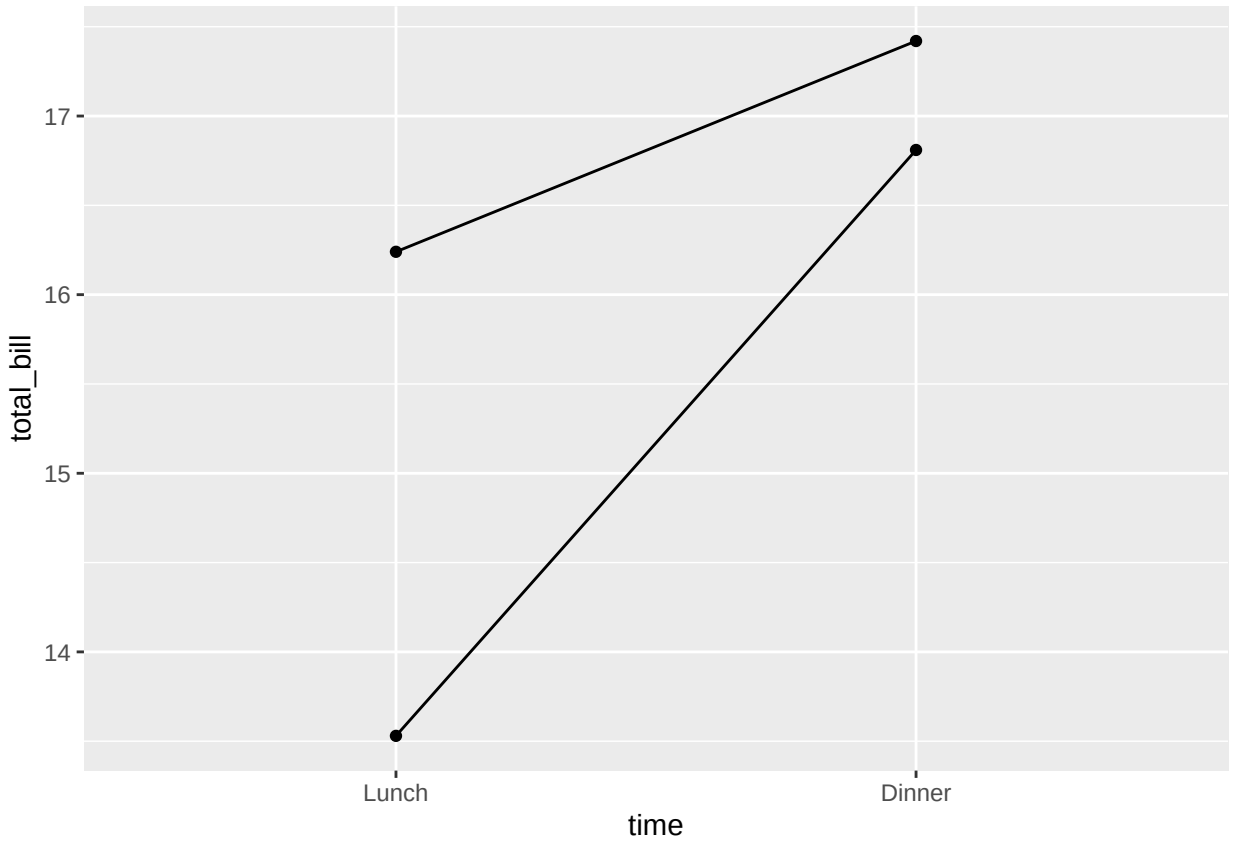

```
# Change colors
ggplot(data=dat1, aes(x=time, y=total_bill, fill=sex)) +
    geom_bar(stat="identity", position=position_dodge(), colour="black") +
    scale_fill_manual(values=c("#999999", "#E69F00"))
```



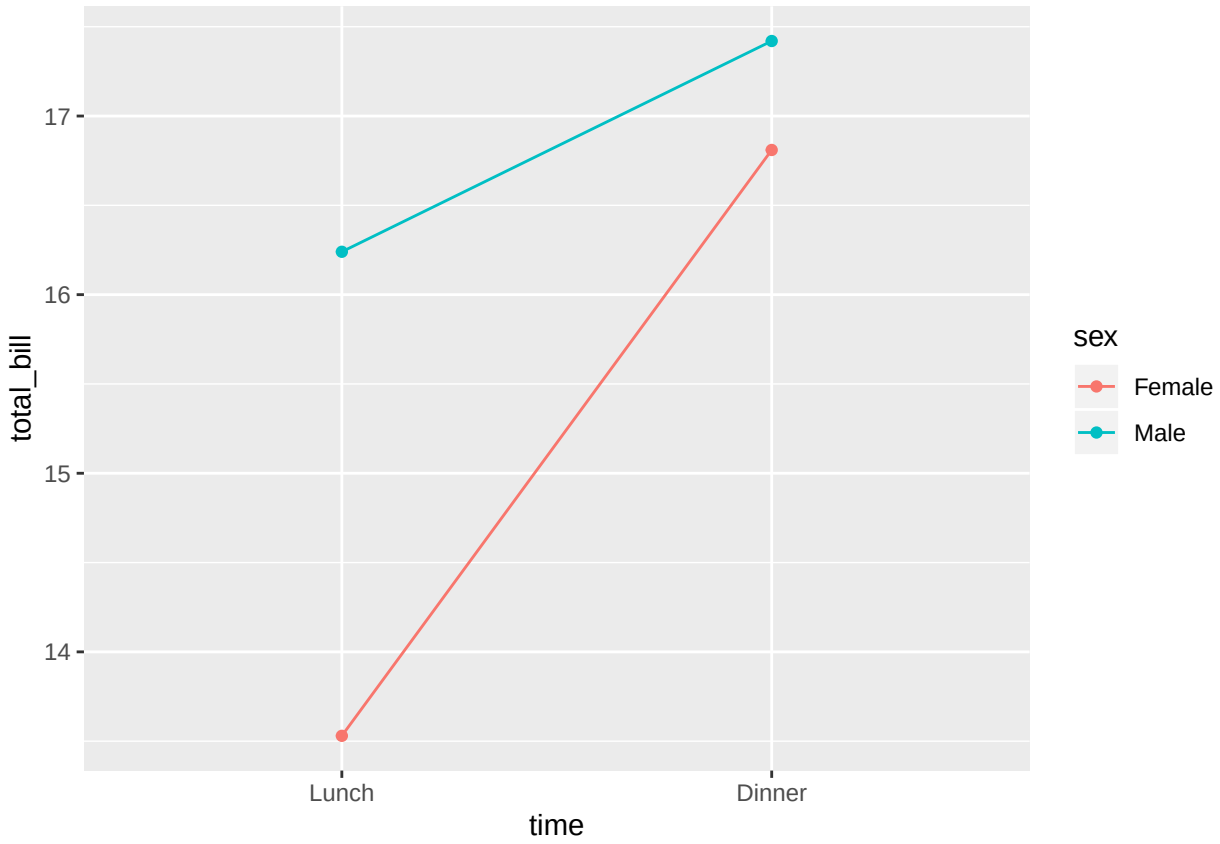

```
# Bar graph, time on x-axis, color fill grouped by sex -- use position_dodge()
ggplot(data=dat1, aes(x=sex, y=total_bill, fill=time)) +
    geom_bar(stat="identity", position=position_dodge(), colour="black")
```



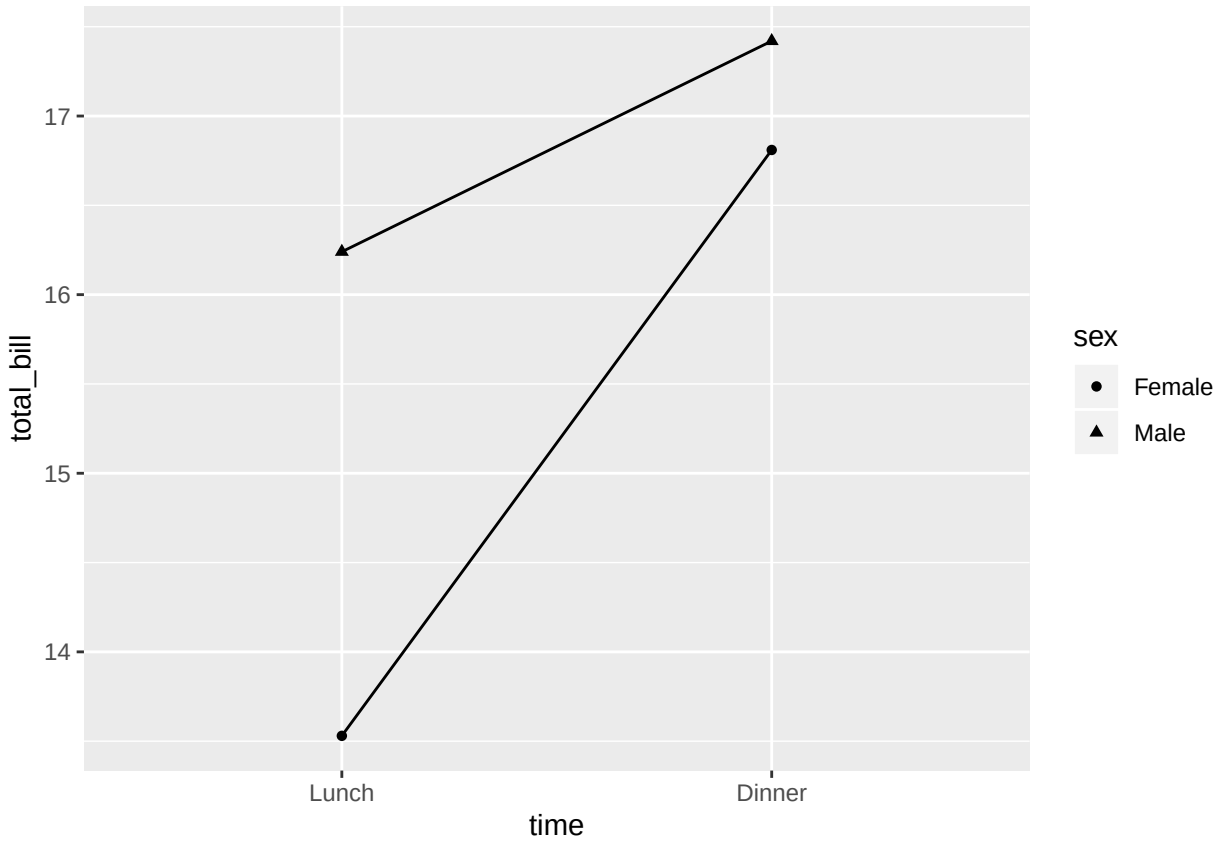

```
# Basic line graph with points
ggplot(data=dat1, aes(x=time, y=total_bill, group=sex)) +
    geom_line() +
    geom_point()
```



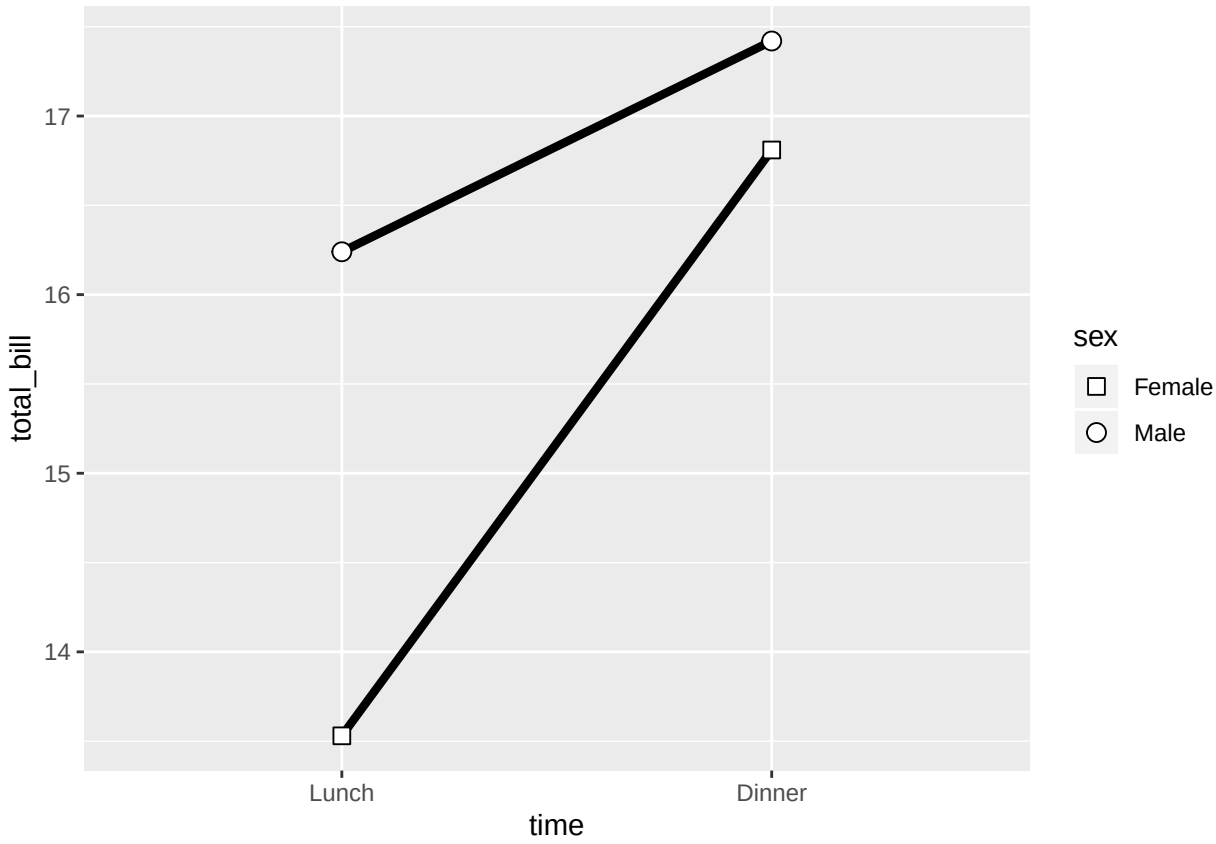

```
# Map sex to color
ggplot(data=dat1, aes(x=time, y=total_bill, group=sex, colour=sex)) +
    geom_line() +
    geom_point()
```



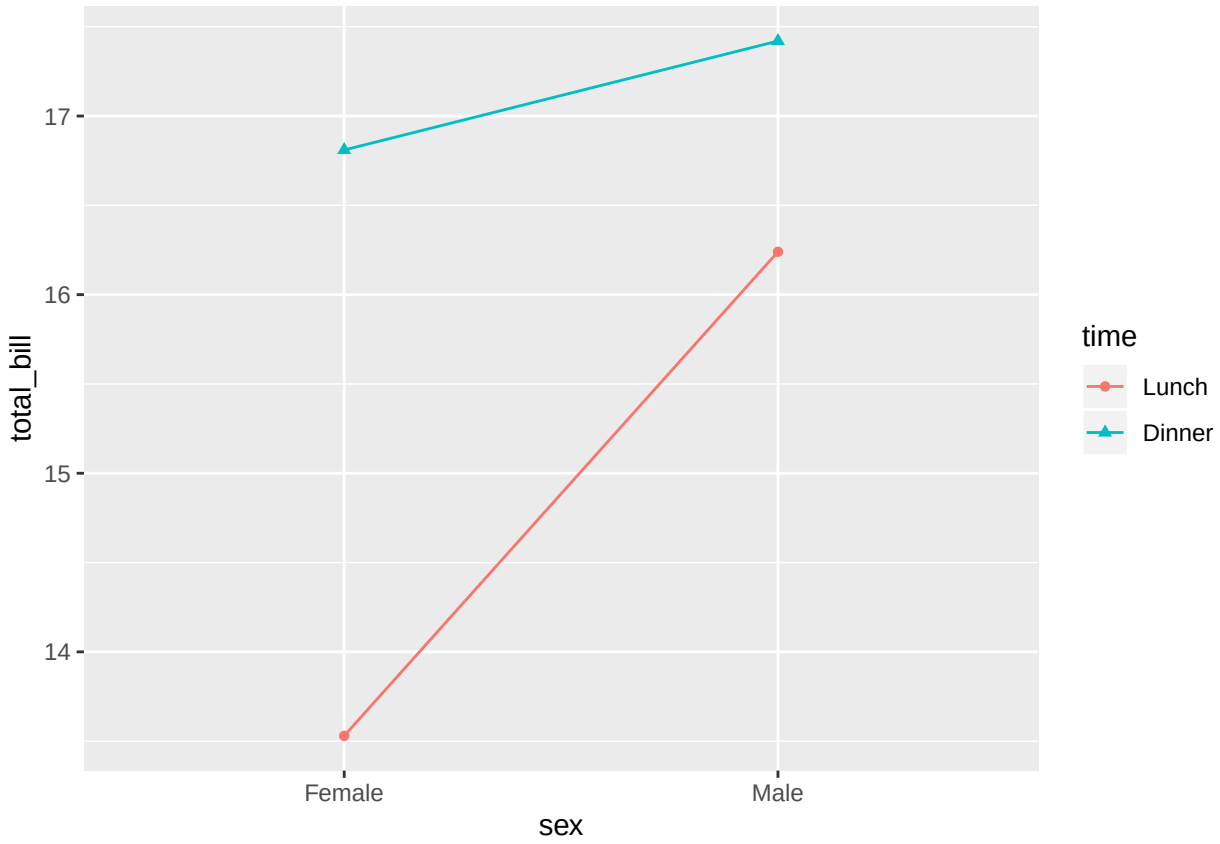

```
# Map sex to different point shape, and use larger points
ggplot(data=dat1, aes(x=time, y=total_bill, group=sex, shape=sex)) +
    geom_line() +
    geom_point()
```




```
# Use thicker lines and larger points, and hollow white-filled points
ggplot(data=dat1, aes(x=time, y=total_bill, group=sex, shape=sex)) +
    geom_line(size=1.5) +
    geom_point(size=3, fill="white") +
    scale_shape_manual(values=c(22,21))
```

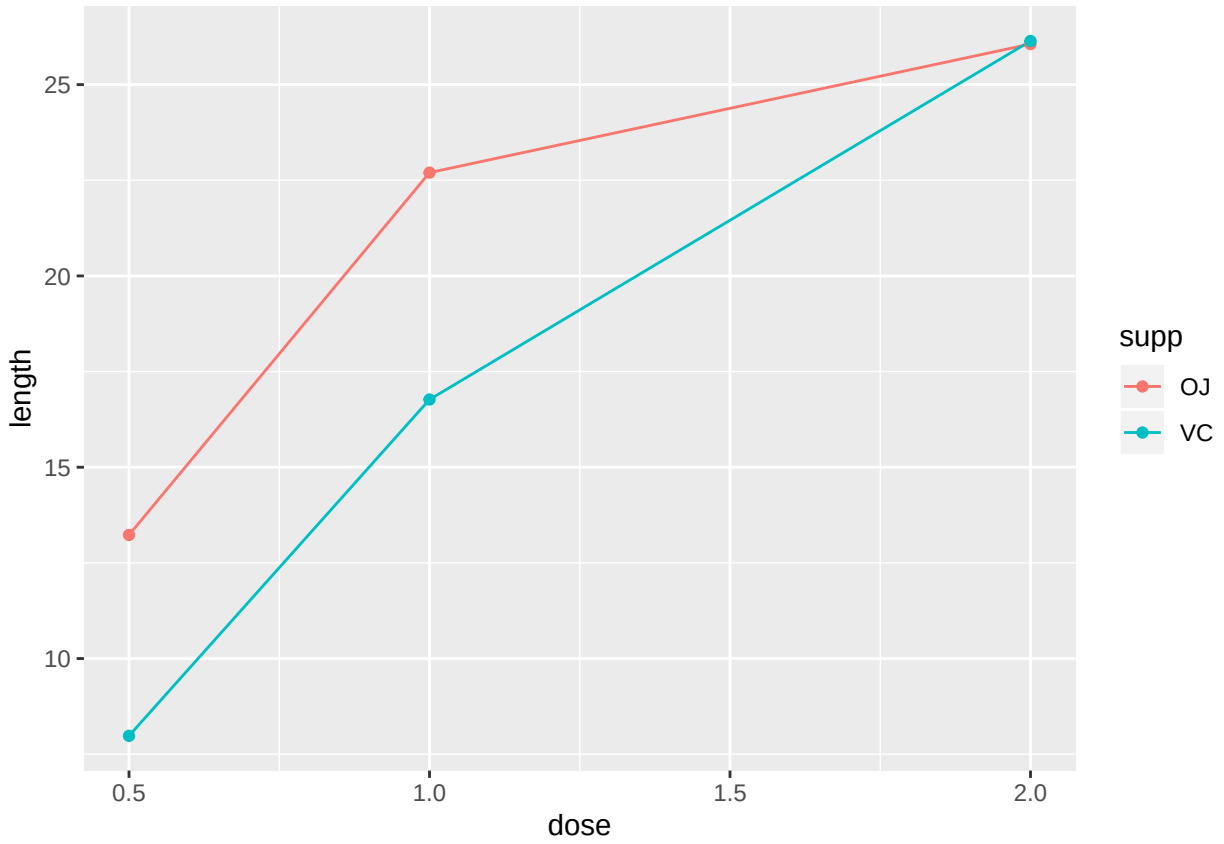



```
ggplot(data=dat1, aes(x=sex, y=total_bill, group=time, shape=time, color=time)) +
    geom_line() +
    geom_point()
```

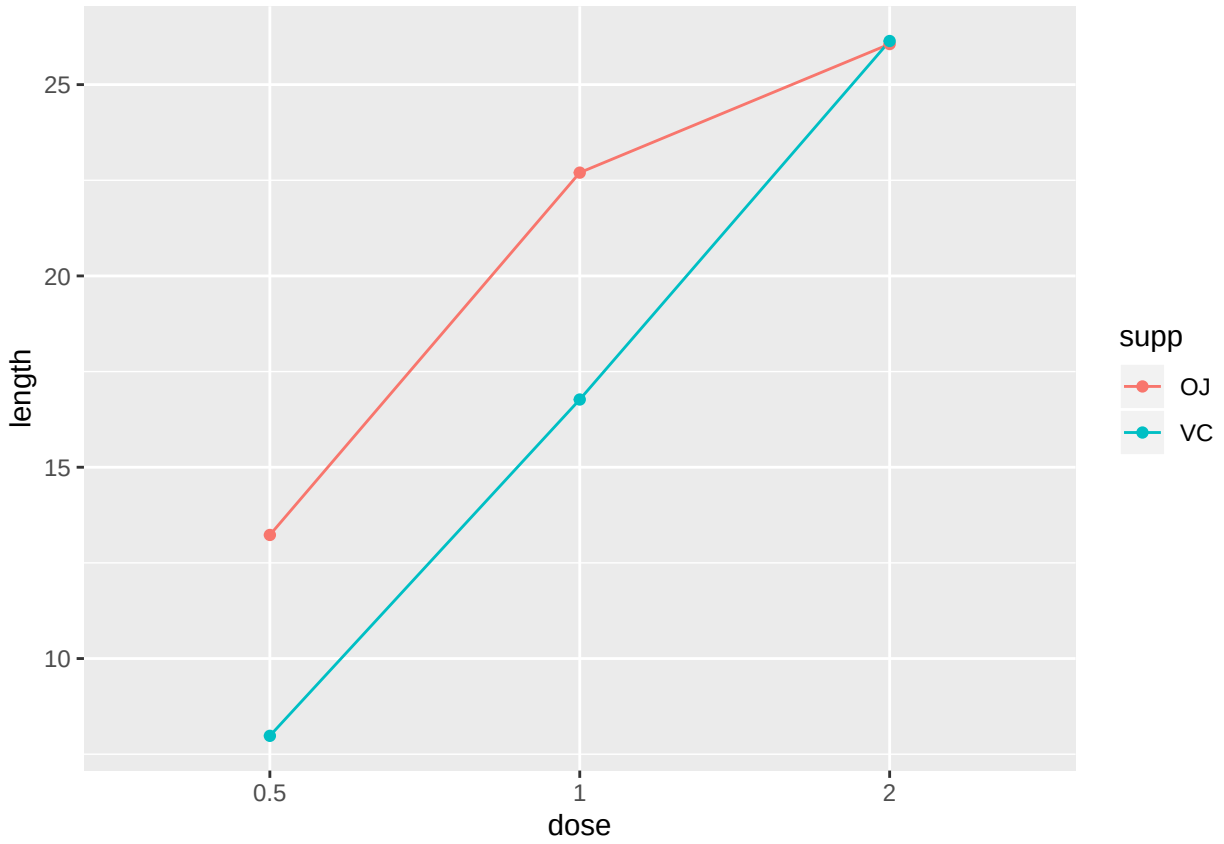



```
datn <- read.table(header=TRUE, text='
supp dose length
  OJ  0.5  13.23
  OJ  1.0  22.70
  OJ  2.0  26.06
  VC  0.5   7.98
  VC  1.0  16.77
  VC  2.0  26.14
')
```

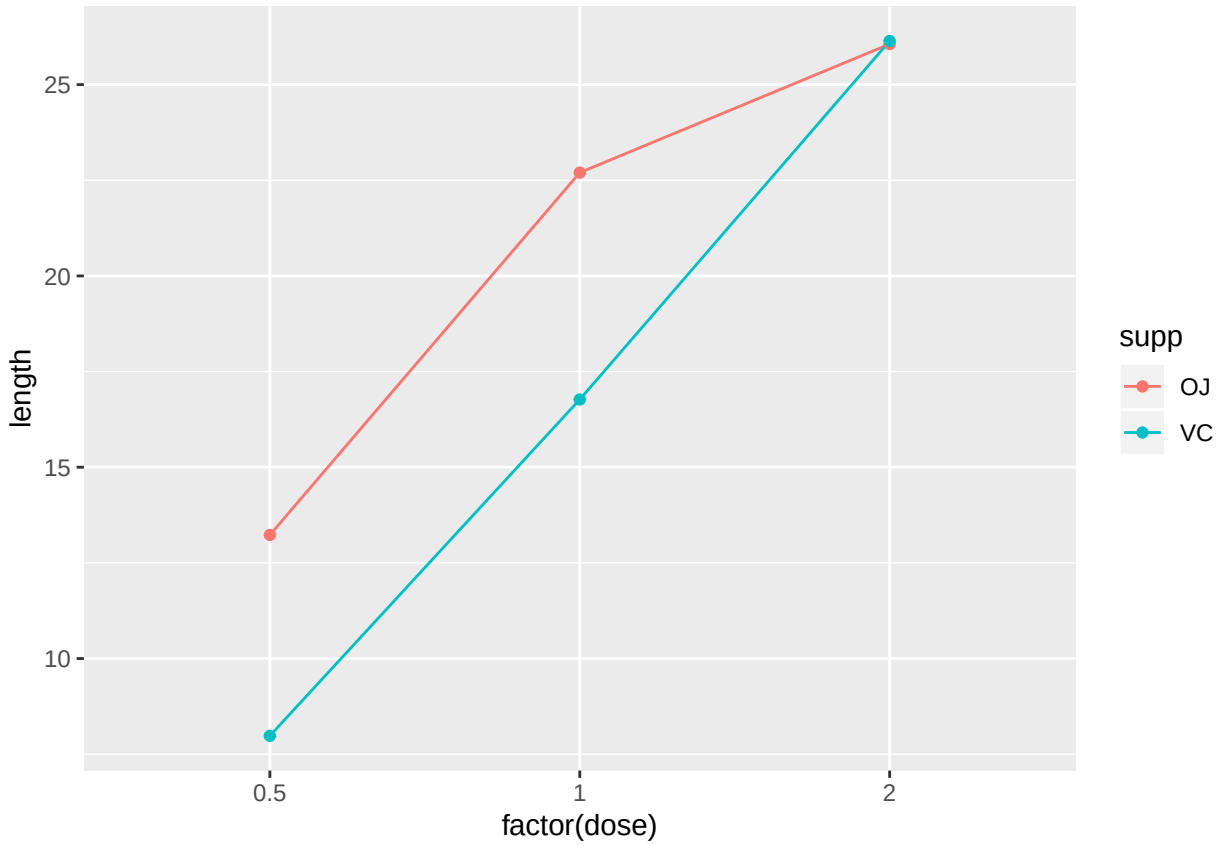
```
ggplot(data=datn, aes(x=dose, y=length, group=supp, colour=supp)) +
    geom_line() +
    geom_point()
```



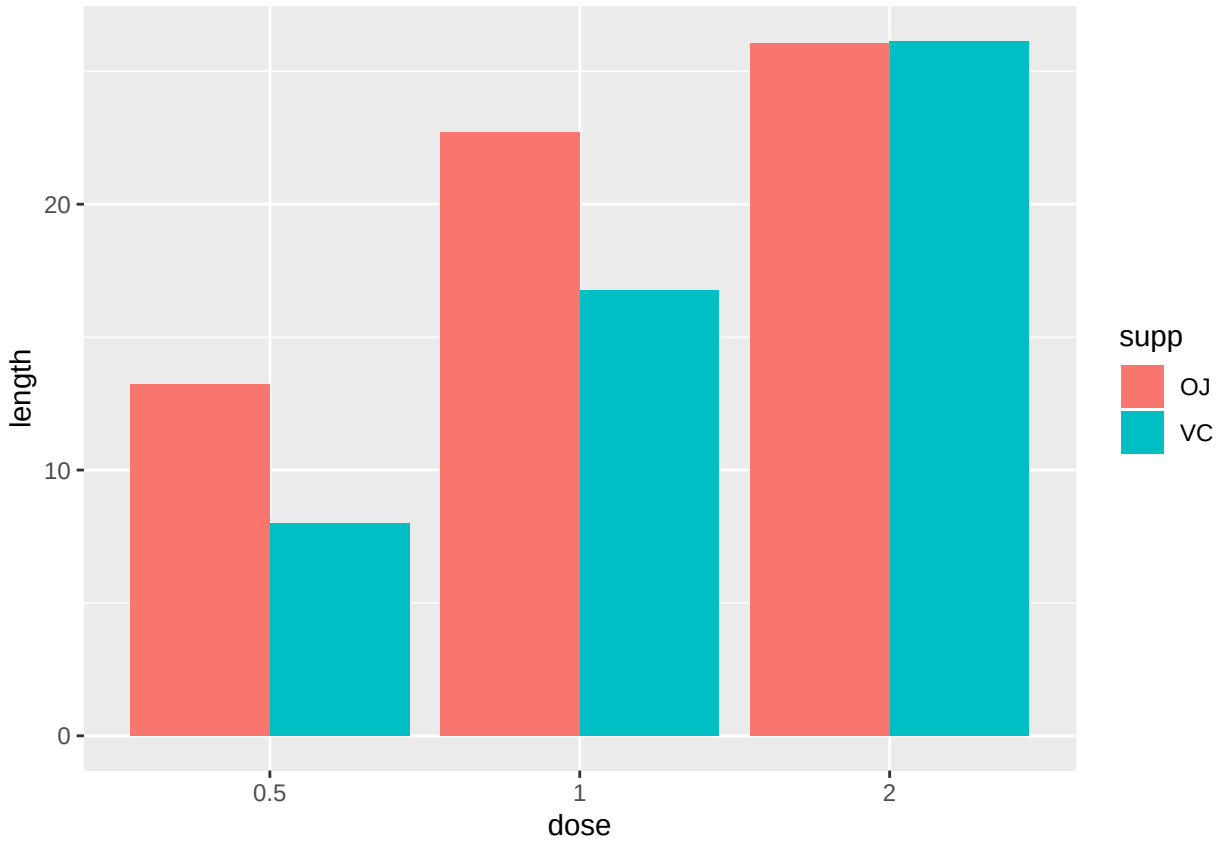

```
# Copy the data frame and convert dose to a factor
datn2 <- datn
datn2$dose <- factor(datn2$dose)
ggplot(data=datn2, aes(x=dose, y=length, group=supp, colour=supp)) +
    geom_line() +
    geom_point()
```



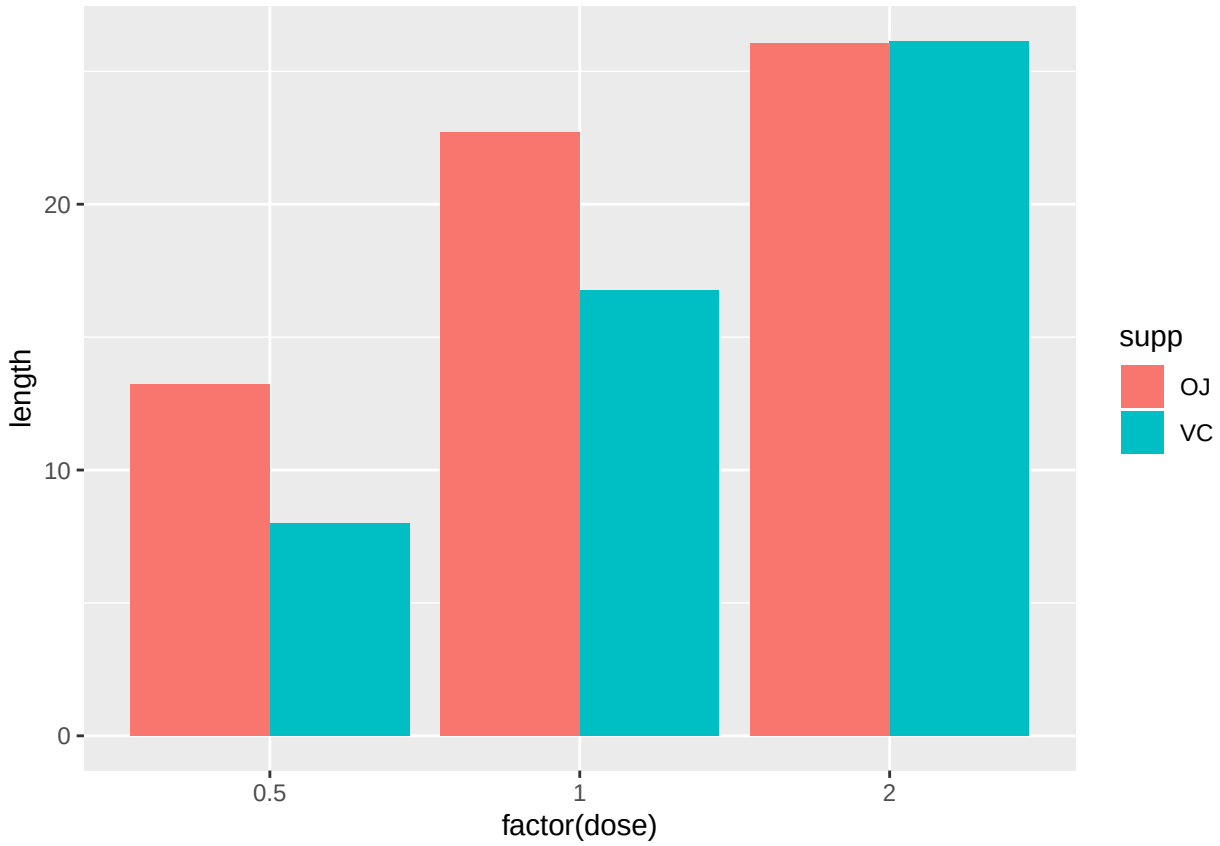

```
# Use the original data frame, but put factor() directly in the plot specification
ggplot(data=datn, aes(x=factor(dose), y=length, group=supp, colour=supp)) +
    geom_line() +
    geom_point()
```




```
# Use datn2 from above
ggplot(data=datn2, aes(x=dose, y=length, fill=supp)) +
    geom_bar(stat="identity", position=position_dodge())
```




```
# Use the original data frame, but put factor() directly in the plot specification
ggplot(data=datn, aes(x=factor(dose), y=length, fill=supp)) +
    geom_bar(stat="identity", position=position_dodge())
```


length
20
10
0
0.5
1
2
factor(dose)
supp
OJ
VC