ggplot2

w

2017年5月7日

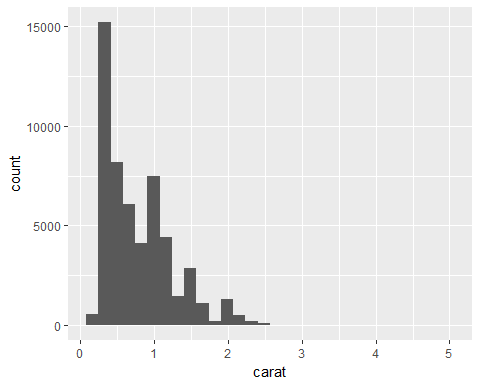
# 基础绘图

## 直方图与密度曲线

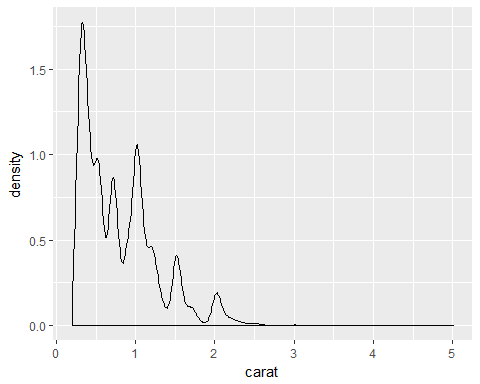
核密度估计是用于估计随机变量概率密度的一种非参数方法

library(ggplot2)  
a<-ggplot(data=diamonds,aes(x=carat))+geom\_histogram()  
a

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



b<-ggplot(data=diamonds)+geom\_density(aes(x=carat))  
b

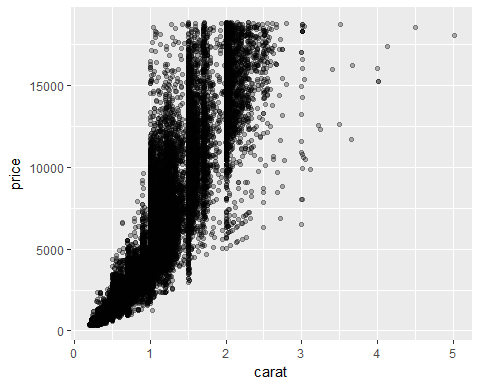


## 散点图

ggplot(diamonds,aes(x=carat,y=price))+geom\_point()

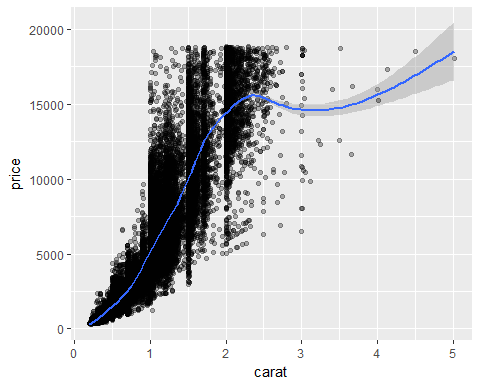


ggplot(diamonds,aes(x=carat,y=price))+geom\_point(alpha=0.3)



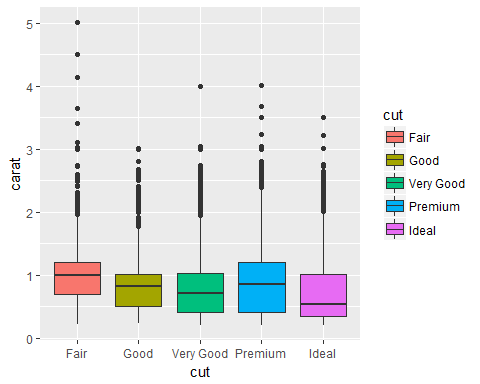
ggplot(diamonds,aes(x=carat,y=price))+geom\_point(alpha=0.3)+geom\_smooth()

## `geom\_smooth()` using method = 'gam'



## 箱线图

ggplot(diamonds,aes(y=carat,x=cut,fill=cut))+geom\_boxplot()

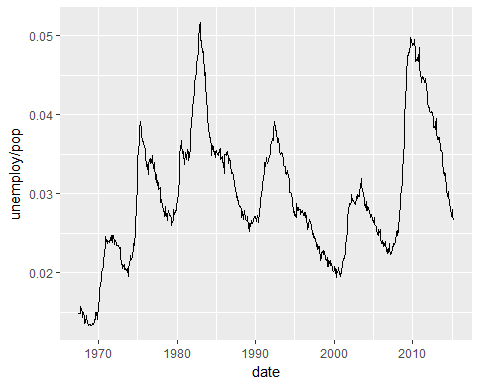


## 曲线图

economics

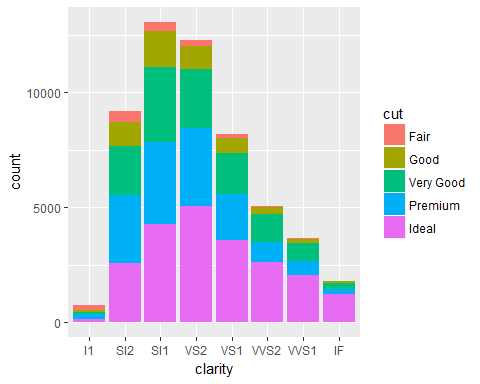
## # A tibble: 574 × 6  
## date pce pop psavert uempmed unemploy  
## <date> <dbl> <int> <dbl> <dbl> <int>  
## 1 1967-07-01 507.4 198712 12.5 4.5 2944  
## 2 1967-08-01 510.5 198911 12.5 4.7 2945  
## 3 1967-09-01 516.3 199113 11.7 4.6 2958  
## 4 1967-10-01 512.9 199311 12.5 4.9 3143  
## 5 1967-11-01 518.1 199498 12.5 4.7 3066  
## 6 1967-12-01 525.8 199657 12.1 4.8 3018  
## 7 1968-01-01 531.5 199808 11.7 5.1 2878  
## 8 1968-02-01 534.2 199920 12.2 4.5 3001  
## 9 1968-03-01 544.9 200056 11.6 4.1 2877  
## 10 1968-04-01 544.6 200208 12.2 4.6 2709  
## # ... with 564 more rows

ggplot(economics,aes(x=date,y=unemploy/pop))+geom\_line()

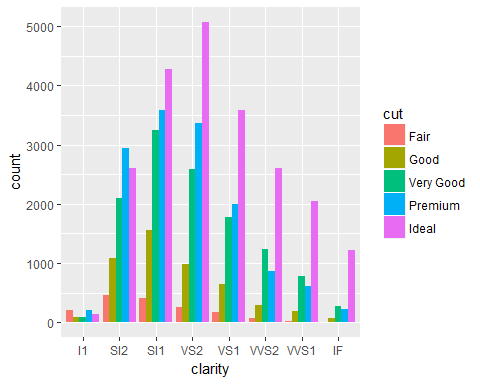


## 条形图

x<-ggplot(diamonds,aes(clarity,fill=cut))  
  
#堆积条形图(将不同切割状态的数据堆积放置)  
  
x1<-x+geom\_bar(position="stack")  
  
  
#簇状条形图（将不同切割状态的数据并列放置）  
  
x2<-x+geom\_bar(position="dodge")  
  
x1

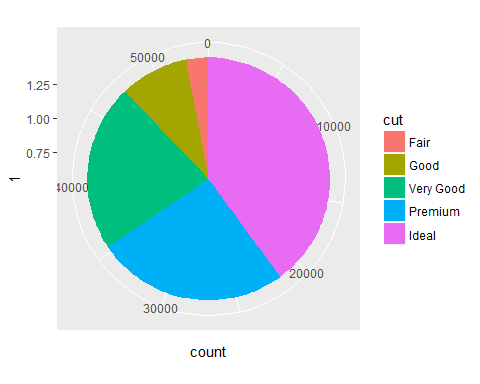


x2

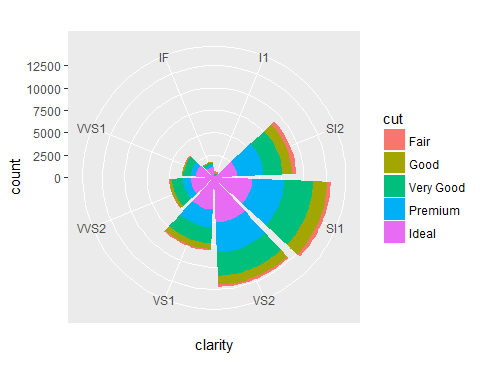


## 饼图

ggplot(diamonds)+geom\_bar(aes(x=1, fill=cut))+coord\_polar(theta='y')

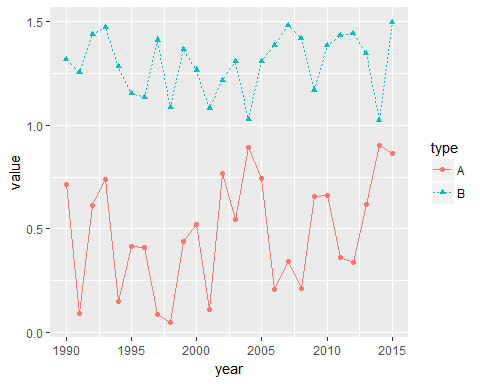


ggplot(diamonds)+geom\_bar(aes(x=clarity, fill=cut))+coord\_polar()



## 折线图

year <- rep(1990:2015, times = 2)  
type <- rep(c('A','B'),each = 26)  
value <- c(runif(26),runif(26, min = 1,max = 1.5))  
df <- data.frame(year = year, type = type, value = value)  
ggplot(df, aes(x = year, y = value, linetype = type, colour = type, shape = type, fill = type))+ geom\_line() + geom\_point()



# 图形进阶

## 直方图与密度曲线

library(ggplot2)  
library(devtools)  
library(easyGgplot2)  
library(grid)  
  
p <- ggplot(mpg,aes(hwy))  
p1<-p + geom\_histogram(position = 'identity',alpha=0.5,aes(y = ..density..,fill =factor(year))) +stat\_density(geom = 'line',aes(colour = factor(year)))  
  
pdf("compare\_identity.pdf")  
  
grid.newpage()  
  
p2<-p + geom\_histogram(alpha=0.5,aes(y = ..density..,fill =factor(year))) +stat\_density(geom = 'line',aes(colour = factor(year)))  
  
pushViewport(viewport(layout=grid.layout(2,1)))  
  
vplayout<-function(x,y)  
 viewport(layout.pos.row=x,layout.pos.col=y)  
  
print(p1,vp=vplayout(1,1))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

print(p2,vp=vplayout(2,1))

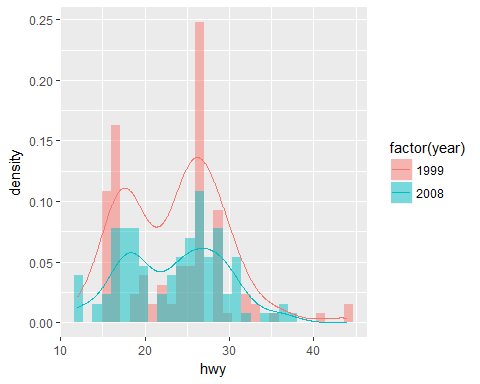
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

dev.off()

## png   
## 2

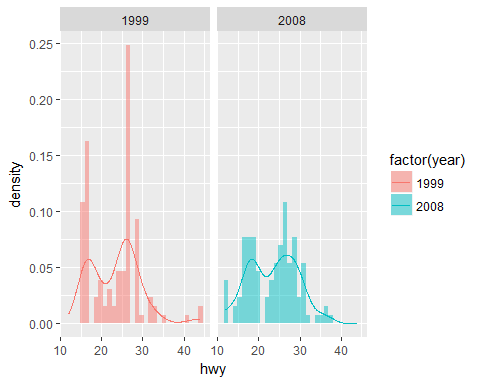
p1

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



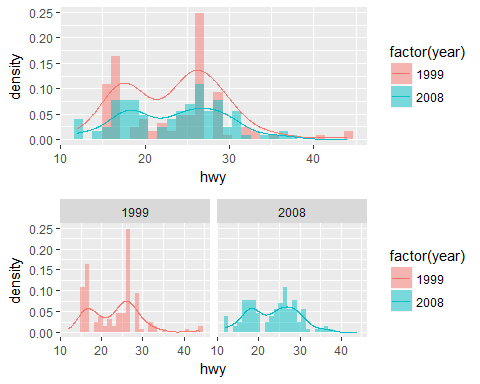
p2<-p1+facet\_wrap(~factor(year))  
  
p2

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggplot2.multiplot(p1,p2,cols=1)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

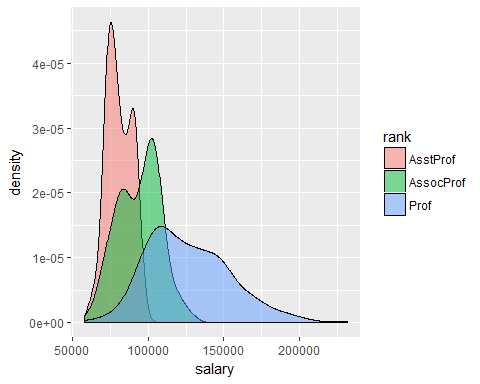


## 密度曲线

data(Salaries,package="car")  
head(Salaries,10)

## rank discipline yrs.since.phd yrs.service sex salary  
## 1 Prof B 19 18 Male 139750  
## 2 Prof B 20 16 Male 173200  
## 3 AsstProf B 4 3 Male 79750  
## 4 Prof B 45 39 Male 115000  
## 5 Prof B 40 41 Male 141500  
## 6 AssocProf B 6 6 Male 97000  
## 7 Prof B 30 23 Male 175000  
## 8 Prof B 45 45 Male 147765  
## 9 Prof B 21 20 Male 119250  
## 10 Prof B 18 18 Female 129000

ggplot(Salaries,aes(x=salary,fill=rank))+geom\_density(alpha=1/2)



## 曲线

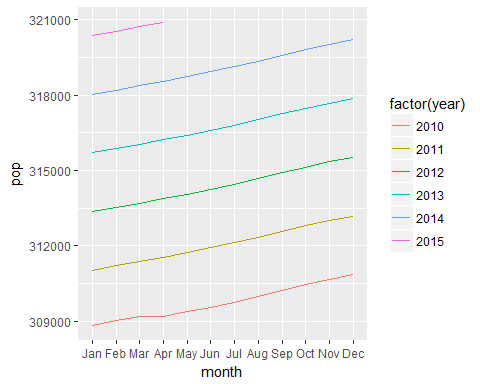
library(lubridate)

## Warning: package 'lubridate' was built under R version 3.3.3

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

economics$year<-year(economics$date)  
  
economics$month<-month(economics$date,label=TRUE)  
  
econ2000<-economics[which(economics$year>=2010),]  
  
g<-ggplot(econ2000,aes(x=month,y=pop))  
  
g<-g+geom\_line(aes(color=factor(year),group=year))  
  
  
g

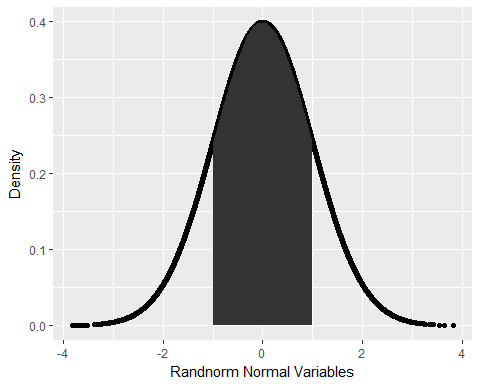


## 多边形的绘制

randnorm<-rnorm(30000)  
randdensity<-dnorm(randnorm)  
p<-ggplot(data.frame(x=randnorm,y=randdensity))+aes(x=x,y=y)+geom\_point()+labs(x="Randnorm Normal Variables",y="Density")  
negpos1seq<-seq(from=-1,to=1,by=0.1)  
negto1<-data.frame(x=negpos1seq,y=dnorm(negpos1seq))  
  
negto1<-rbind(c(min(negto1),0),negto1,c(max(negto1$x),0))  
head(negto1)

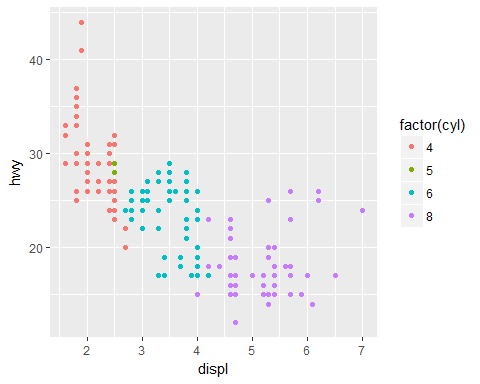
## x y  
## 1 -1.0 0.0000000  
## 2 -1.0 0.2419707  
## 3 -0.9 0.2660852  
## 4 -0.8 0.2896916  
## 5 -0.7 0.3122539  
## 6 -0.6 0.3332246

p+geom\_polygon(data=negto1,aes(x=x,y=y))

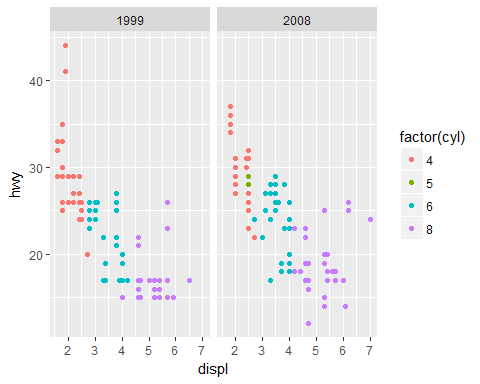


## 散点图

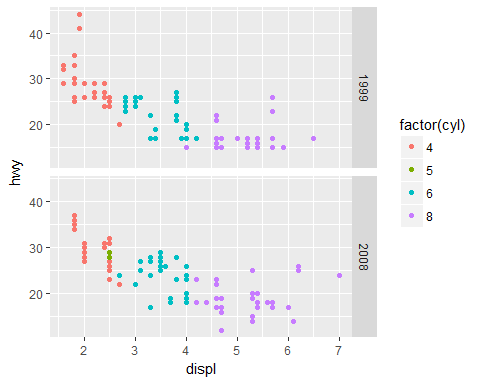
p <- ggplot(data=mpg,aes(x=displ,y=hwy,colour=factor(cyl)))  
  
p1<-p + geom\_point()   
p1



p1+facet\_grid(.~year) #列项展示

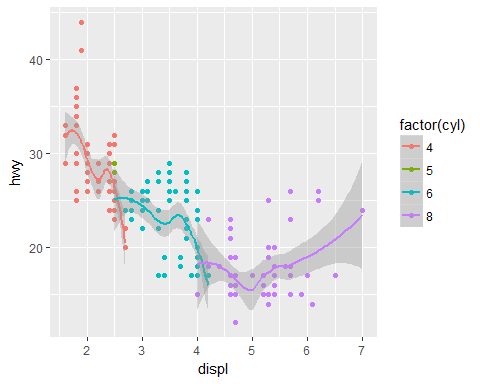


p1+facet\_grid(year~.) #行项展示



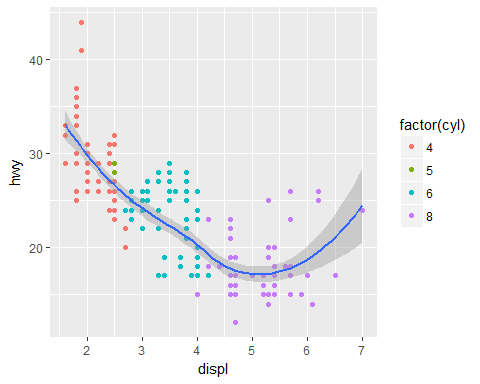
p1+ geom\_smooth()

## `geom\_smooth()` using method = 'loess'



g<- ggplot(mpg,aes(x=displ,y=hwy))  
g+ geom\_point(aes(colour=factor(cyl))) + geom\_smooth()

## `geom\_smooth()` using method = 'loess'



## 气泡散点图

bmp(filename="output.bmp")  
p <- ggplot(mpg, aes(x=cty,y=hwy))  
p + geom\_point(aes(colour=factor(year),size=displ),alpha=0.5,position = "jitter")+ stat\_smooth()+scale\_color\_manual(values =c('blue2','red4'))+  
scale\_size\_continuous(range = c(4, 10))+coord\_cartesian(xlim = c(15, 25),ylim=c(15,40))

## `geom\_smooth()` using method = 'loess'

dev.off()

## png   
## 2

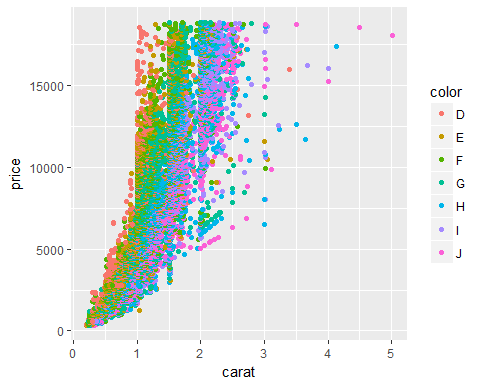
## 主题

library(grid)  
require(ggthemes)

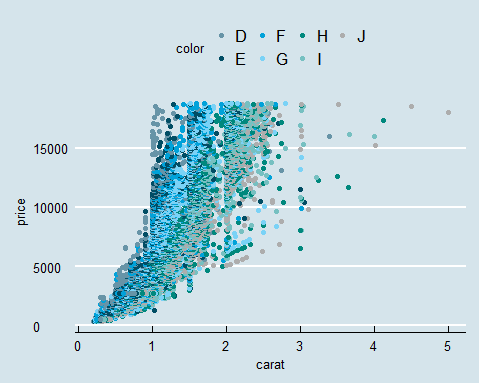
## Loading required package: ggthemes

## Warning: package 'ggthemes' was built under R version 3.3.3

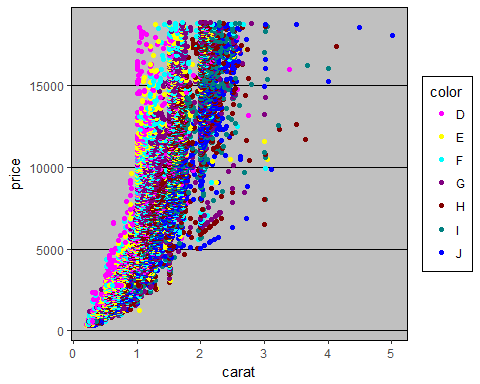
g<-ggplot(diamonds,aes(x=carat,y=price))+geom\_point(aes(color=color))  
g



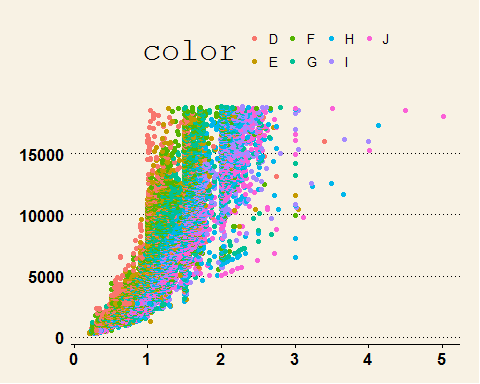
g1<-g+theme\_economist()+scale\_colour\_economist()  
  
g2<-g+theme\_excel()+scale\_colour\_excel()  
  
g3<-g+theme\_wsj()  
  
g1



g2



g3

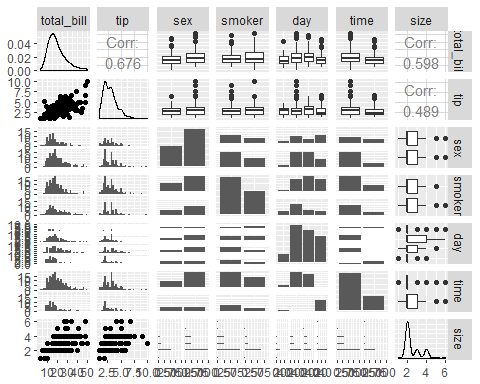


data(tips,package="reshape2")  
head(tips)

## 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

GGally::ggpairs(tips)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggsave("graph.pdf", width=10, height=10,limitsize = FALSE)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.