DS\_HW3

RE6124019\_吳明軒

2023-10-29

# Chapter 2  
  
# 2.2  
library(ggplot2)  
mpg

## # A tibble: 234 × 11  
## manufacturer model displ year cyl trans drv cty hwy fl class  
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>  
## 1 audi a4 1.8 1999 4 auto… f 18 29 p comp…  
## 2 audi a4 1.8 1999 4 manu… f 21 29 p comp…  
## 3 audi a4 2 2008 4 manu… f 20 31 p comp…  
## 4 audi a4 2 2008 4 auto… f 21 30 p comp…  
## 5 audi a4 2.8 1999 6 auto… f 16 26 p comp…  
## 6 audi a4 2.8 1999 6 manu… f 18 26 p comp…  
## 7 audi a4 3.1 2008 6 auto… f 18 27 p comp…  
## 8 audi a4 quattro 1.8 1999 4 manu… 4 18 26 p comp…  
## 9 audi a4 quattro 1.8 1999 4 auto… 4 16 25 p comp…  
## 10 audi a4 quattro 2 2008 4 manu… 4 20 28 p comp…  
## # ℹ 224 more rows

# 2.3  
ggplot(mpg, aes(x = displ, y = hwy)) +  
 geom\_point()



ggplot(mpg, aes(displ, hwy)) +  
 geom\_point()



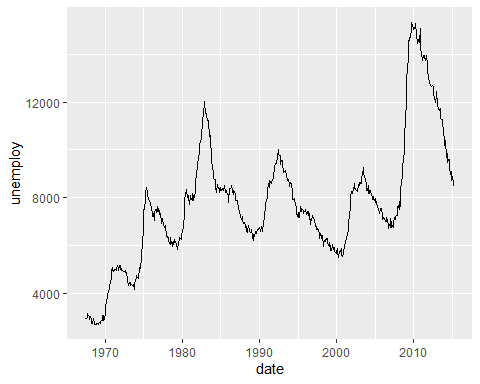
ggplot(mpg, aes(cty, hwy)) +   
 geom\_point()



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

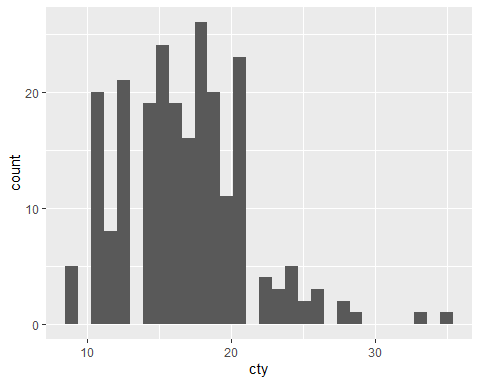


ggplot(economics, aes(date, unemploy)) +  
 geom\_line()

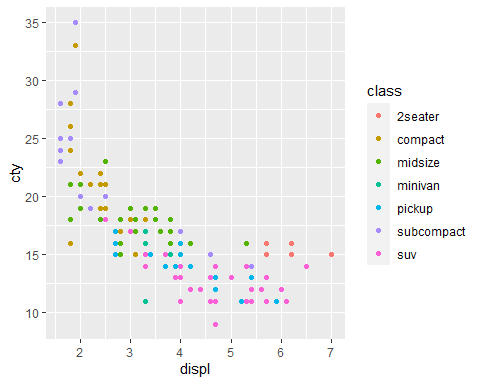


ggplot(mpg, aes(cty)) +  
 geom\_histogram()

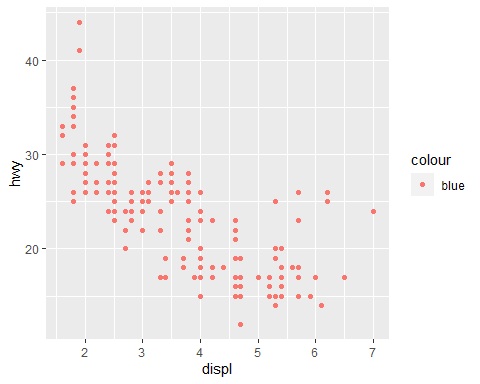
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



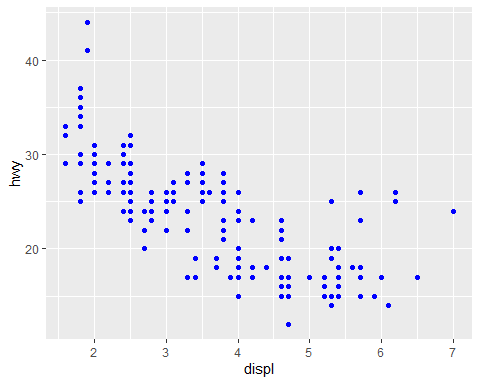
# 2.4  
ggplot(mpg, aes(displ, cty, colour = class)) +  
 geom\_point()



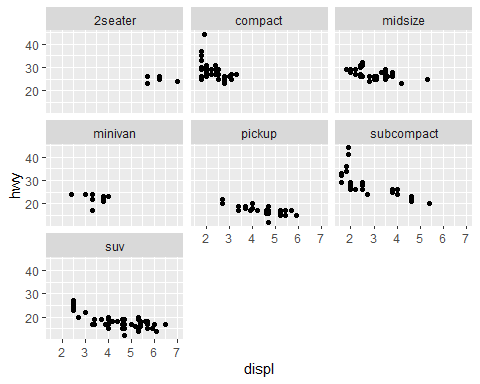
ggplot(mpg, aes(displ, hwy)) +   
 geom\_point(aes(colour = "blue"))



ggplot(mpg, aes(displ, hwy)) +  
 geom\_point(colour = "blue")

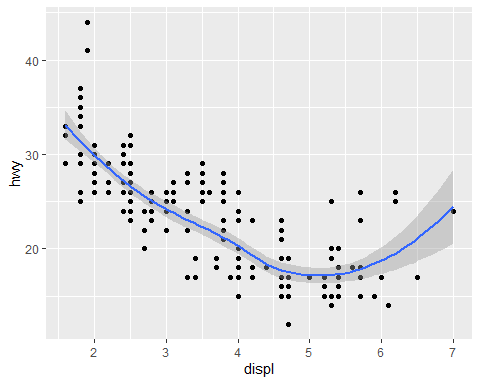


# 2.5  
ggplot(mpg, aes(displ, hwy)) +  
 geom\_point() +  
 facet\_wrap(~class)



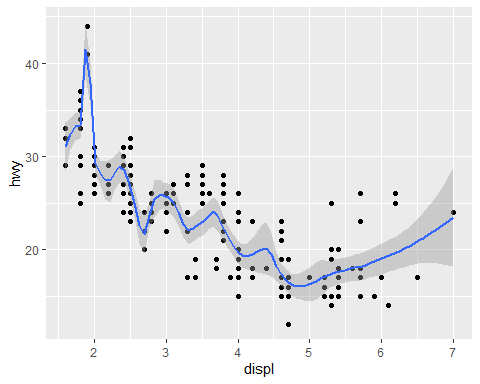
# 2.6  
ggplot(mpg, aes(displ, hwy)) +  
 geom\_point() +  
 geom\_smooth()

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'



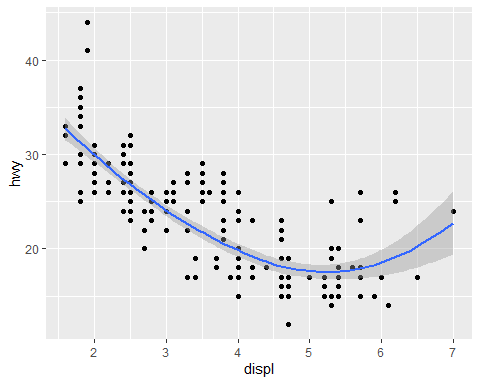
ggplot(mpg, aes(displ, hwy)) +  
 geom\_point() +  
 geom\_smooth(span = 0.2)

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'



ggplot(mpg, aes(displ, hwy)) +  
 geom\_point() +  
 geom\_smooth(span = 1)

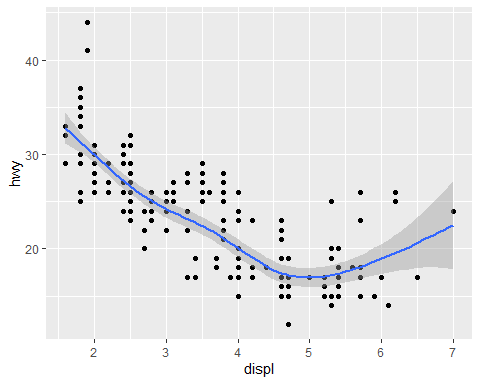
## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'



library(mgcv)

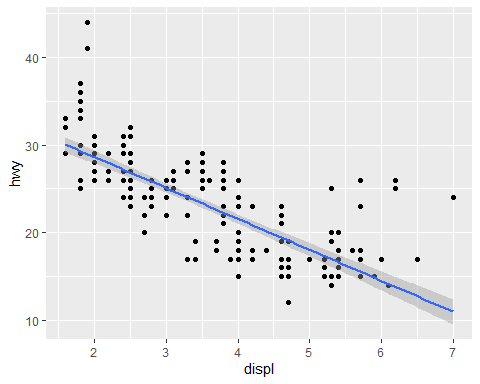
## Loading required package: nlme  
## This is mgcv 1.8-42. For overview type 'help("mgcv-package")'.

ggplot(mpg, aes(displ, hwy)) +  
 geom\_point() +  
 geom\_smooth(method = "gam", formula = y ~ s(x))

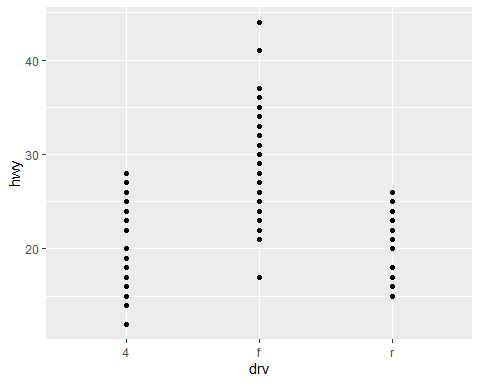


ggplot(mpg, aes(displ, hwy)) +  
 geom\_point() +   
 geom\_smooth(method = "lm")

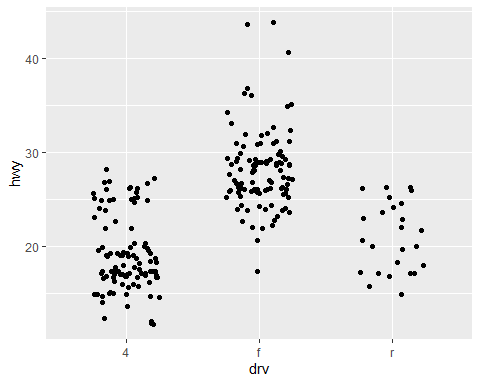
## `geom\_smooth()` using formula = 'y ~ x'



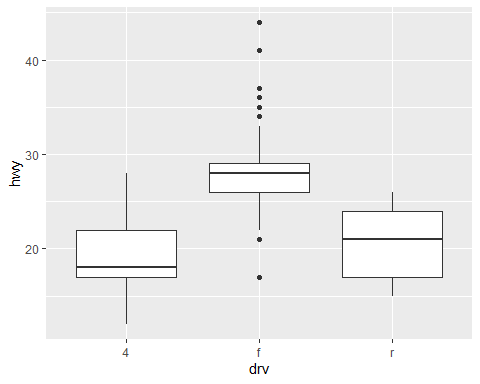
ggplot(mpg, aes(drv, hwy)) +  
 geom\_point()



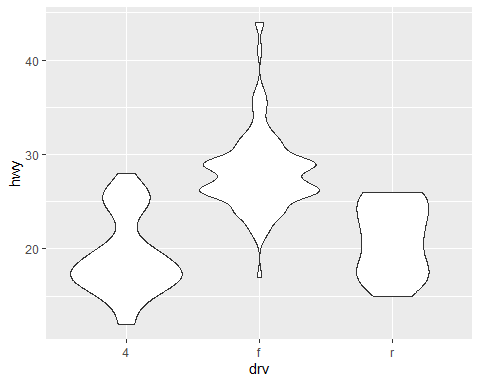
ggplot(mpg, aes(drv, hwy)) +   
 geom\_jitter(width = 0.25)



ggplot(mpg, aes(drv, hwy)) +  
 geom\_boxplot()

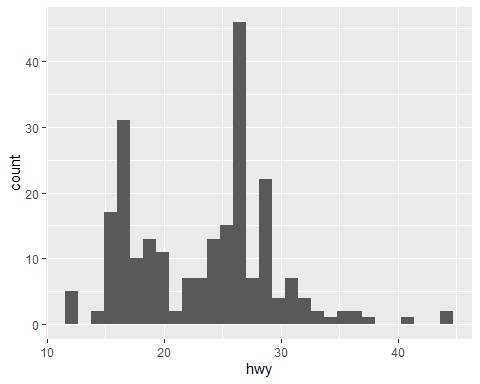


ggplot(mpg, aes(drv, hwy)) +  
 geom\_violin()



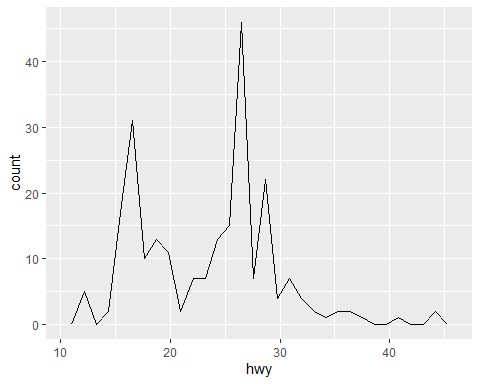
ggplot(mpg, aes(hwy)) +  
 geom\_histogram()

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

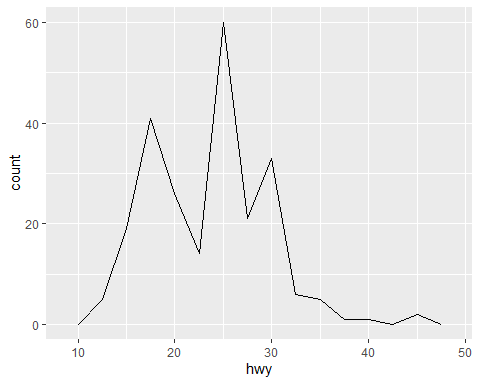


ggplot(mpg, aes(hwy)) +  
 geom\_freqpoly()

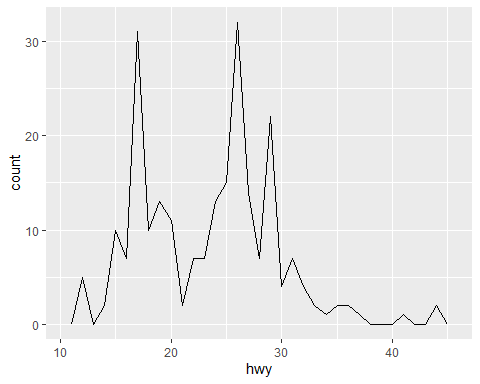
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



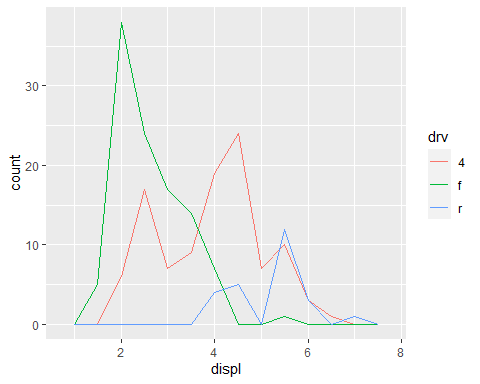
ggplot(mpg, aes(hwy)) +  
 geom\_freqpoly(binwidth = 2.5)



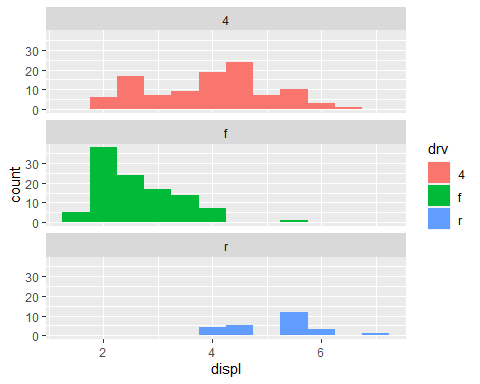
ggplot(mpg, aes(hwy)) +  
 geom\_freqpoly(binwidth = 1)



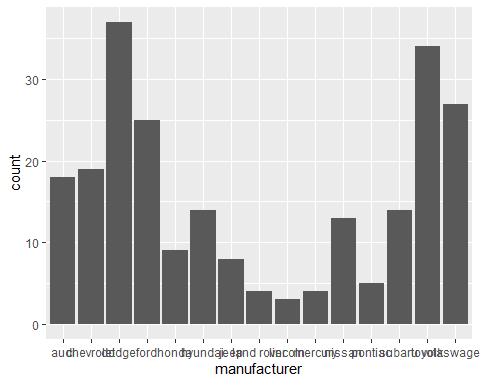
ggplot(mpg, aes(displ, colour = drv)) +  
 geom\_freqpoly(binwidth = 0.5)



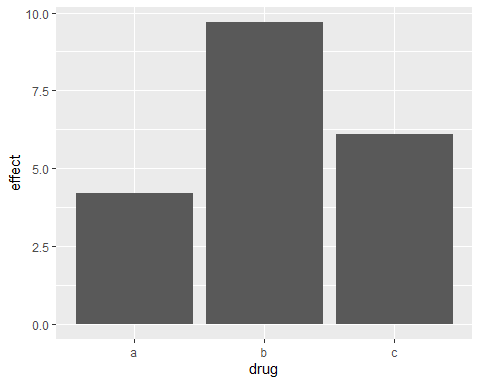
ggplot(mpg, aes(displ, fill = drv)) +  
 geom\_histogram(binwidth = 0.5) +  
 facet\_wrap(~drv, ncol = 1)



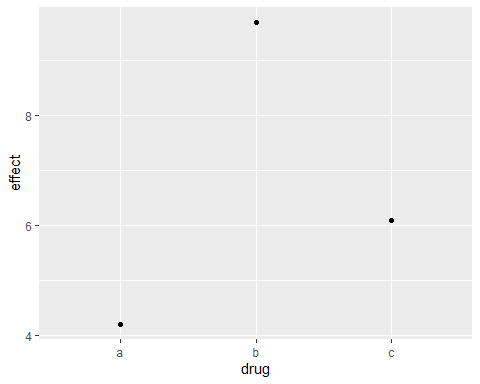
ggplot(mpg, aes(manufacturer)) +  
 geom\_bar()



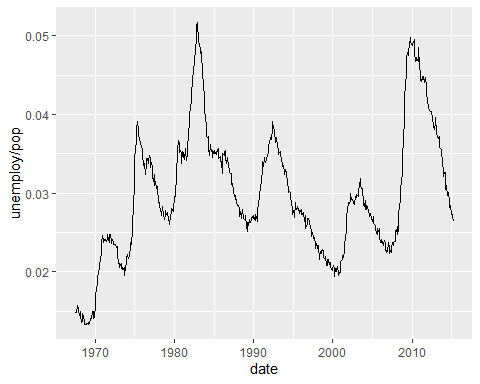
drugs <- data.frame(  
 drug = c("a", "b", "c"),  
 effect = c(4.2, 9.7, 6.1)  
)  
  
ggplot(drugs, aes(drug, effect)) +   
 geom\_bar(stat = "identity")



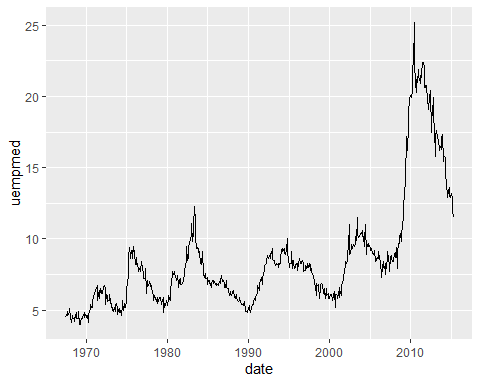
ggplot(drugs, aes(drug, effect)) +  
 geom\_point()



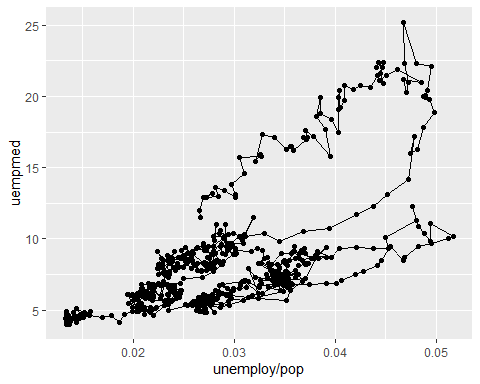
ggplot(economics, aes(date, unemploy / pop)) +  
 geom\_line()



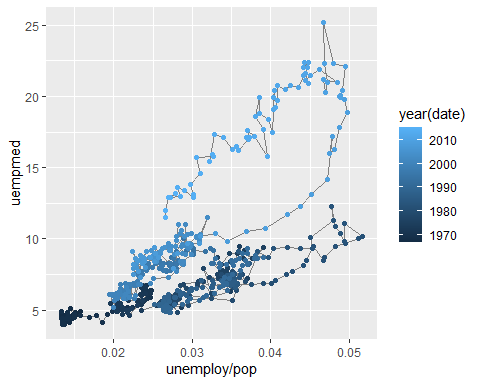
ggplot(economics, aes(date, uempmed)) +  
 geom\_line()



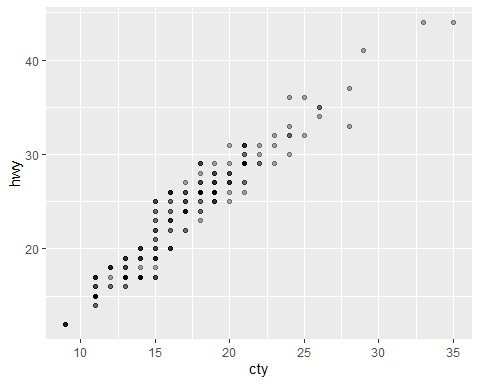
ggplot(economics, aes(unemploy / pop, uempmed)) +  
 geom\_path() +  
 geom\_point()



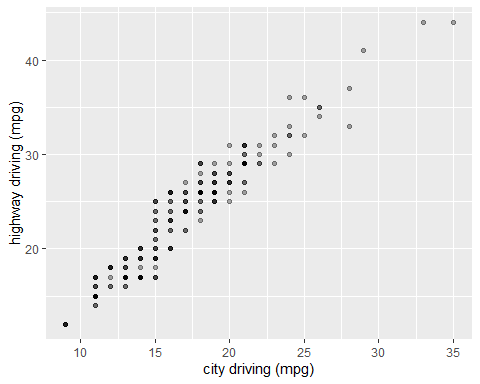
year <- function(x) as.POSIXlt(x)$year + 1900  
ggplot(economics, aes(unemploy / pop, uempmed)) +  
 geom\_path(colour = "grey50") +  
 geom\_point(aes(colour = year(date)))



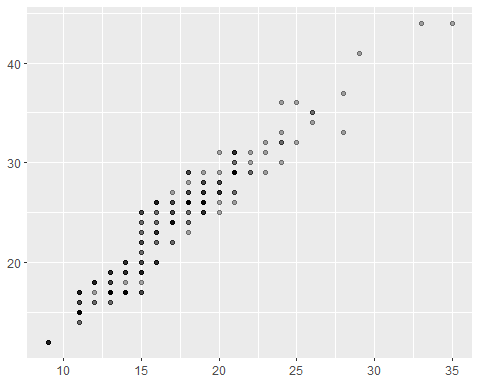
# 2.7  
ggplot(mpg, aes(cty, hwy)) +  
 geom\_point(alpha = 1 / 3)



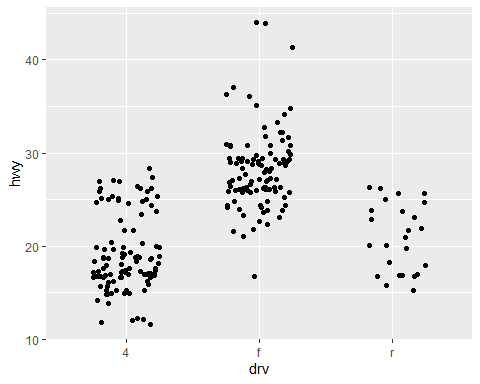
ggplot(mpg, aes(cty, hwy)) +  
 geom\_point(alpha = 1 / 3) +  
 xlab("city driving (mpg)") +  
 ylab("highway driving (mpg)")



ggplot(mpg, aes(cty, hwy)) +  
 geom\_point(alpha = 1 / 3) +  
 xlab(NULL) +  
 ylab(NULL)

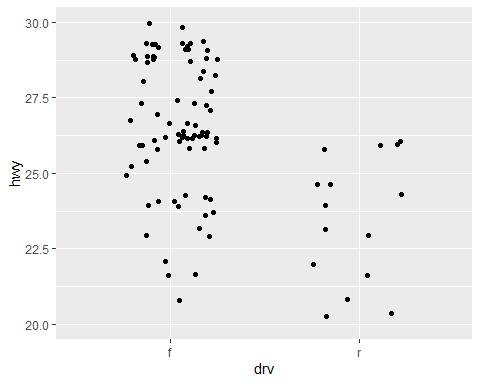


ggplot(mpg, aes(drv, hwy)) +  
 geom\_jitter(width = 0.25)

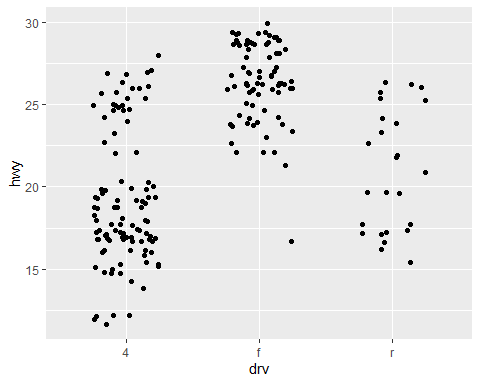


ggplot(mpg, aes(drv, hwy)) +  
 geom\_jitter(width = 0.25) +  
 xlim("f", "r") +  
 ylim(20, 30)

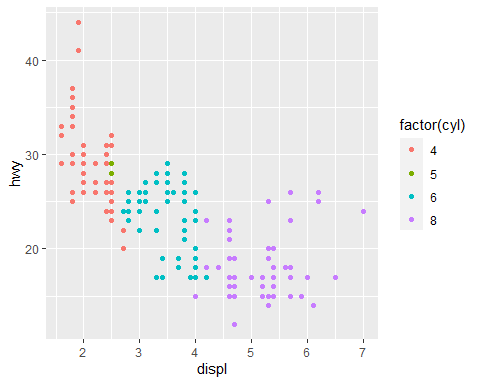
## Warning: Removed 138 rows containing missing values (`geom\_point()`).



ggplot(mpg, aes(drv, hwy)) +  
 geom\_jitter(width = 0.25, na.rm = TRUE) +   
 ylim(NA, 30)



# 2.8  
p <- ggplot(mpg, aes(displ, hwy, colour = factor(cyl))) +  
 geom\_point()  
  
print(p)



ggsave("plot.png", width = 5, height = 5)  
  
summary(p)

## data: manufacturer, model, displ, year, cyl, trans, drv, cty, hwy, fl,  
## class [234x11]  
## mapping: x = ~displ, y = ~hwy, colour = ~factor(cyl)  
## faceting: <ggproto object: Class FacetNull, Facet, gg>  
## compute\_layout: function  
## draw\_back: function  
## draw\_front: function  
## draw\_labels: function  
## draw\_panels: function  
## finish\_data: function  
## init\_scales: function  
## map\_data: function  
## params: list  
## setup\_data: function  
## setup\_params: function  
## shrink: TRUE  
## train\_scales: function  
## vars: function  
## super: <ggproto object: Class FacetNull, Facet, gg>  
## -----------------------------------  
## geom\_point: na.rm = FALSE  
## stat\_identity: na.rm = FALSE  
## position\_identity

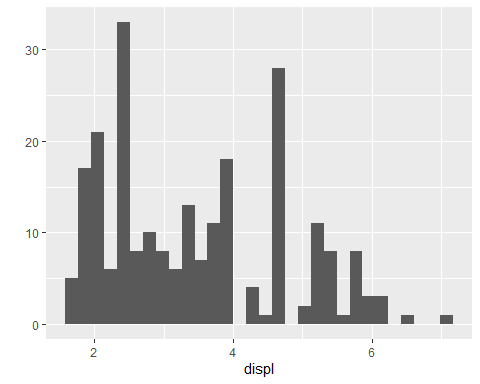
saveRDS(p, "plot.rds")  
q <- readRDS("plot.rds")  
  
# 2.9  
qplot(displ, hwy, data = mpg)

## Warning: `qplot()` was deprecated in ggplot2 3.4.0.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

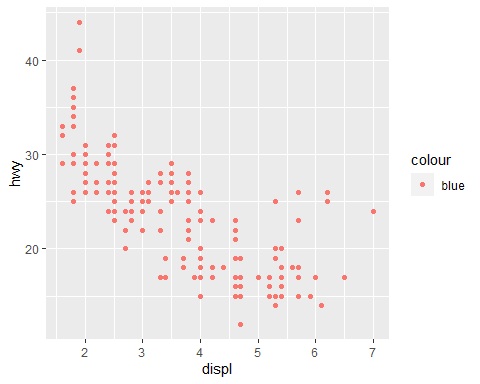


qplot(displ, data = mpg)

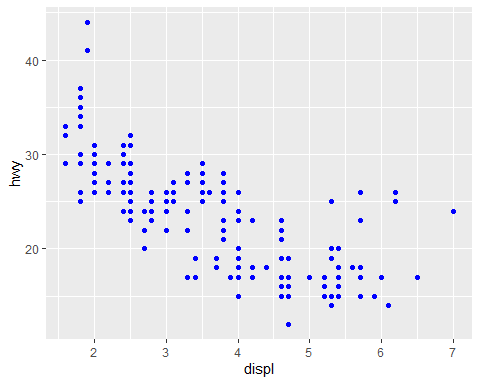
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



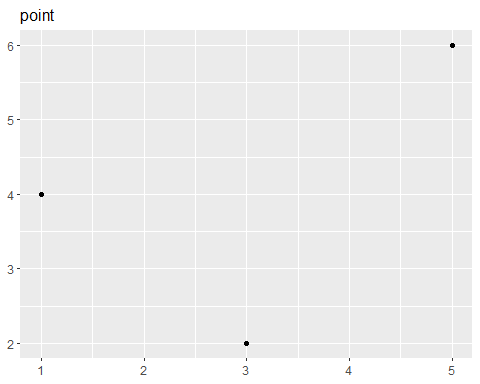
qplot(displ, hwy, data = mpg, colour = "blue")



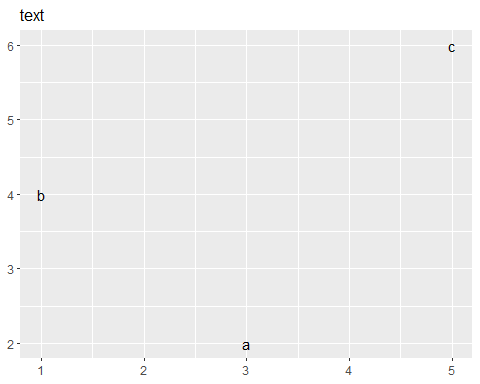
qplot(displ, hwy, data = mpg, colour = I("blue"))



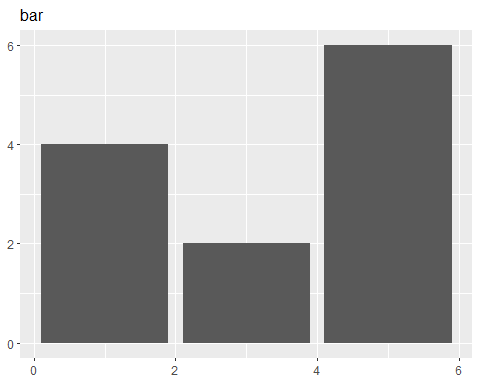
# Chapter 3  
  
library(ggplot2)  
  
#3.2  
df <- data.frame(  
 x = c(3, 1, 5),  
 y = c(2, 4, 6),  
 label = c("a","b","c")  
)  
p <- ggplot(df, aes(x, y, label = label)) +  
 labs(x = NULL, y = NULL) + # Hide axis label   
 theme(plot.title = element\_text(size = 12)) # Shrink plot title  
  
p + geom\_point() + ggtitle("point")



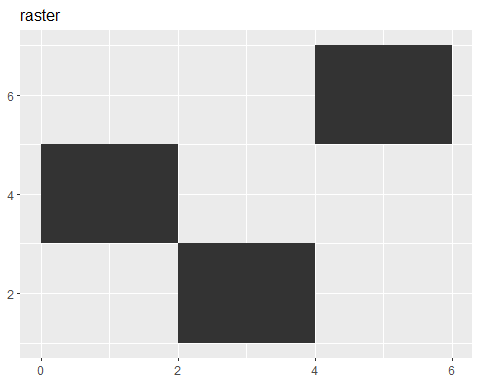
p + geom\_text() + ggtitle("text")



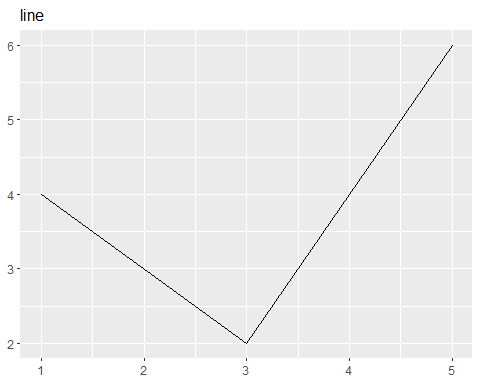
p + geom\_bar(stat = "identity") + ggtitle("bar")



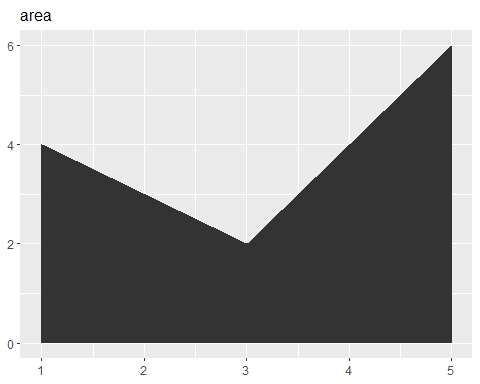
p + geom\_tile() + ggtitle("raster")



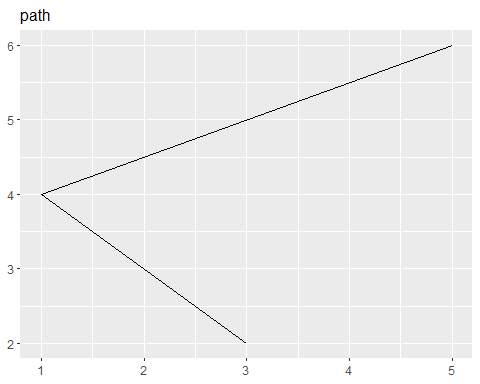
p + geom\_line() + ggtitle("line")



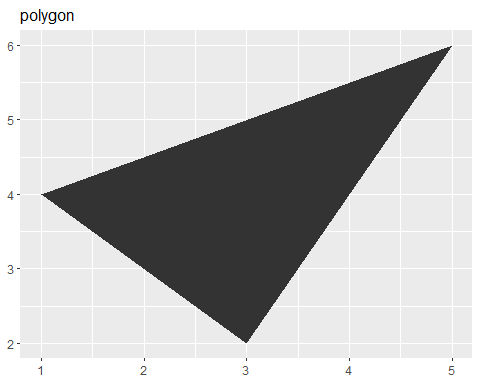
p + geom\_area() + ggtitle("area")



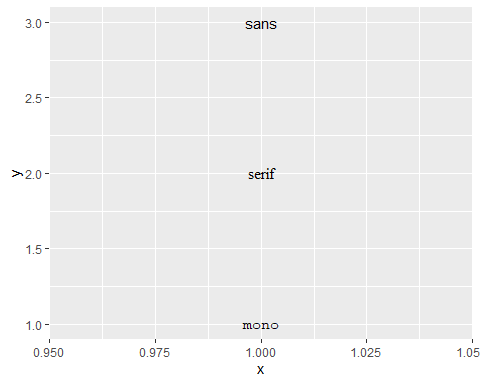
p + geom\_path() + ggtitle("path")



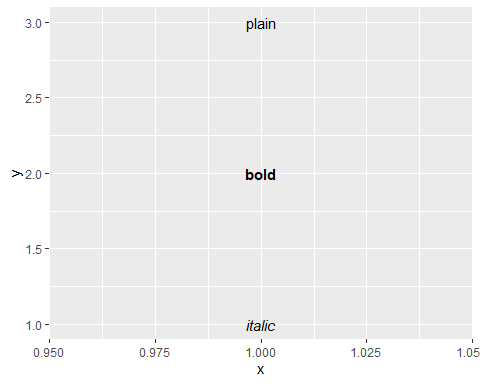
p + geom\_polygon() + ggtitle("polygon")



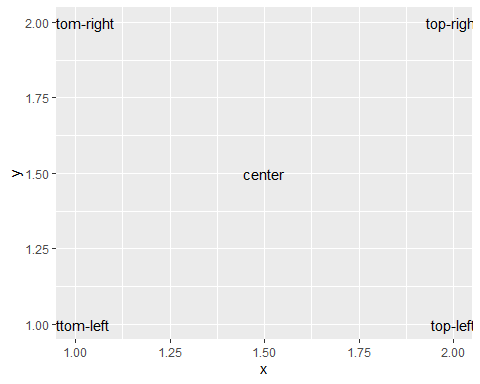
# 3.3  
df <- data.frame(x = 1, y = 3:1, family = c("sans", "serif", "mono"))  
ggplot(df, aes(x, y)) +  
 geom\_text(aes(label = family, family = family))



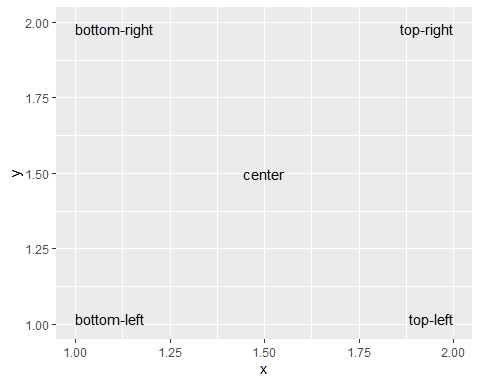
df <- data.frame(x = 1, y = 3:1, face = c("plain", "bold", "italic"))  
ggplot(df, aes(x, y)) +  
 geom\_text(aes(label = face, fontface = face))



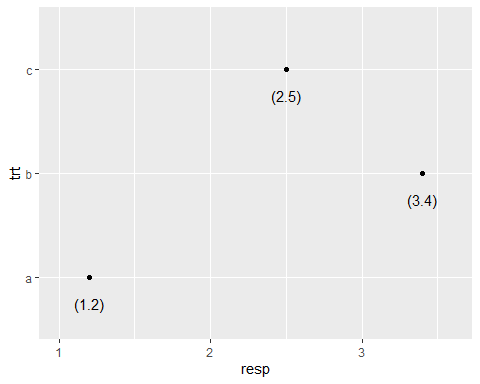
df <- data.frame(  
 x = c(1, 1, 2, 2, 1.5),  
 y = c(1, 2, 1, 2, 1.5),  
 text = c(  
 "bottom-left", "bottom-right",   
 "top-left", "top-right", "center"  
 )  
)  
  
ggplot(df, aes(x, y)) +  
 geom\_text(aes(label = text))



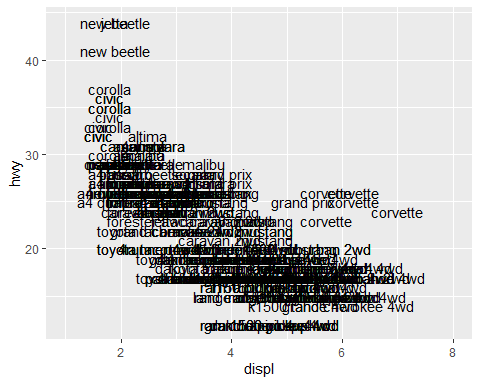
ggplot(df, aes(x, y)) +  
 geom\_text(aes(label = text), vjust = "inward", hjust = "inward")



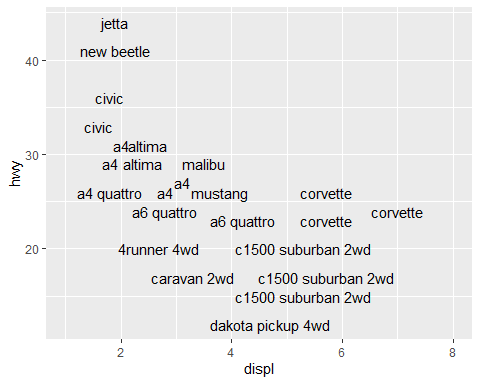
df <- data.frame(trt = c("a", "b", "c"), resp = c(1.2, 3.4, 2.5))  
ggplot(df, aes(resp, trt)) +  
 geom\_point() +  
 geom\_text(aes(label = paste0("(", resp, ")")), nudge\_y = -0.25) +  
 xlim(1, 3.6)



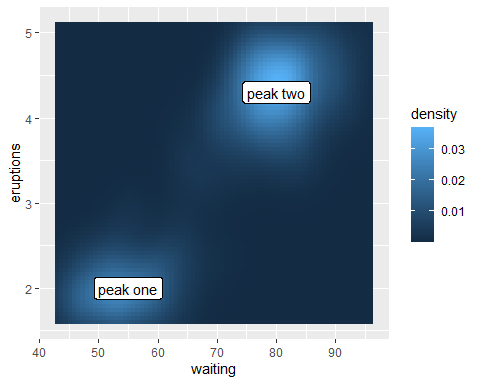
ggplot(mpg, aes(displ, hwy)) +  
 geom\_text(aes(label = model)) +  
 xlim(1, 8)



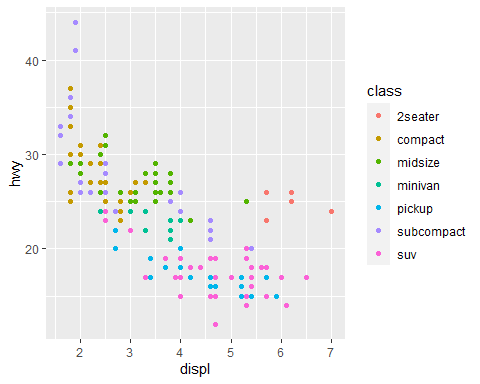
ggplot(mpg, aes(displ, hwy)) +  
 geom\_text(aes(label = model), check\_overlap = TRUE) +   
 xlim(1, 8)



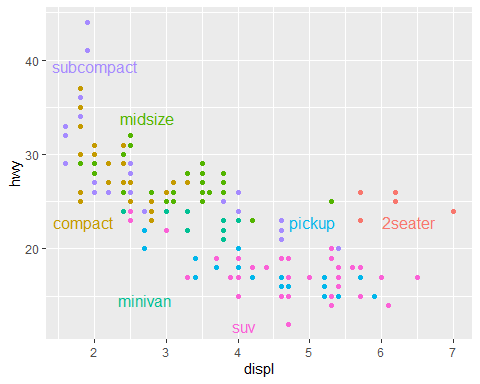
label <- data.frame(  
 waiting = c(55, 80),  
 eruptions = c(2, 4.3),  
 label = c("peak one", "peak two")   
)  
  
ggplot(faithfuld, aes(waiting, eruptions)) +  
 geom\_tile(aes(fill = density)) +  
 geom\_label(data = label, aes(label = label))



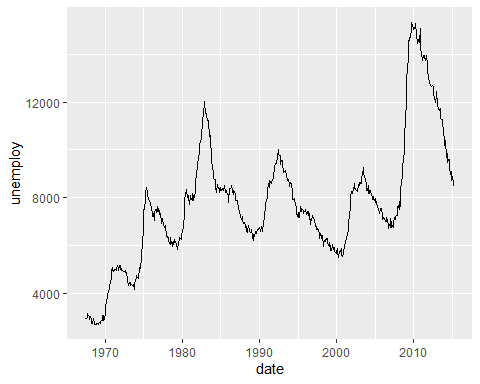
ggplot(mpg, aes(displ, hwy, colour = class)) +  
 geom\_point()



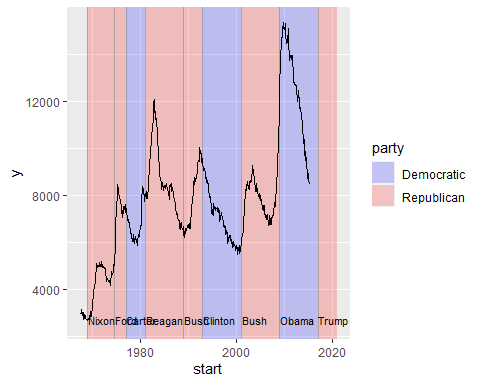
ggplot(mpg, aes(displ, hwy, colour = class)) +  
 geom\_point(show.legend = FALSE) +  
 directlabels::geom\_dl(aes(label = class), method = "smart.grid")



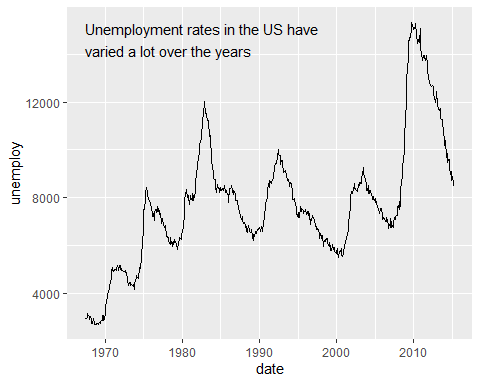
# 3.4  
ggplot(economics, aes(date, unemploy)) +  
 geom\_line()



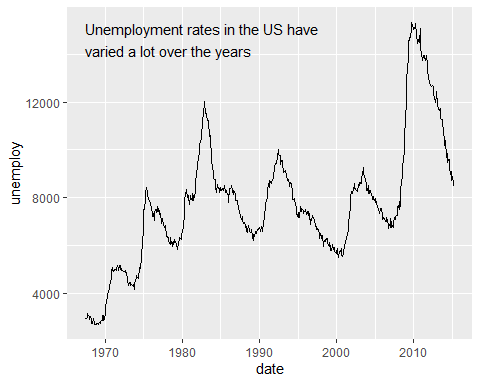
presidential <- subset(presidential, start > economics$date[1])  
  
ggplot(economics) +  
 geom\_rect(  
 aes(xmin = start, xmax = end, fill = party),  
 ymin = -Inf, ymax = Inf, alpha = 0.2,  
 data = presidential  
 ) +  
 geom\_vline(  
 aes(xintercept = as.numeric(start)),  
 data = presidential,  
 colour = "grey50", alpha = 0.5  
 ) +  
 geom\_text(  
 aes(x = start, y = 2500, label = name),   
 data = presidential,  
 size = 3, vjust = 0, hjust = 0, nudge\_x = 50  
 ) +  
 geom\_line(aes(date, unemploy)) +  
 scale\_fill\_manual(values = c("blue", "red"))



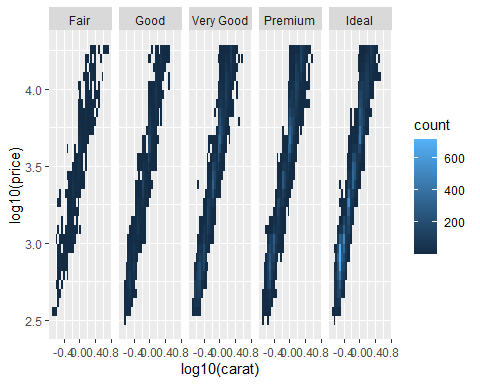
yrng <- range(economics$unemploy)  
xrng <- range(economics$date)  
caption <- paste(strwrap("Unemployment rates in the US have varied a lot over the years", 40), collapse = "\n")  
  
ggplot(economics, aes(date, unemploy)) +  
 geom\_line() +  
 geom\_text(  
 aes(x, y, label = caption),  
 data = data.frame(x = xrng[1], y = yrng[2], caption = caption),   
 hjust = 0, vjust = 1, size = 4  
 )



ggplot(economics, aes(date, unemploy)) +  
 geom\_line() +  
 annotate("text", x = xrng[1], y = yrng[2], label = caption,   
 hjust = 0, vjust = 1, size = 4  
 )

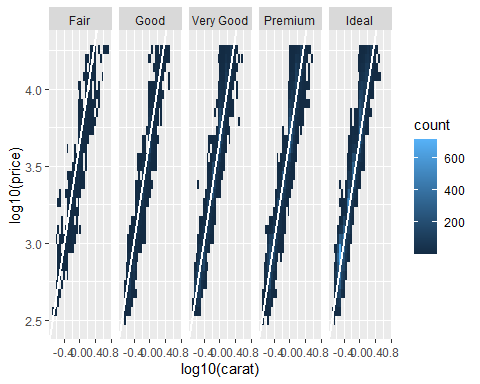


ggplot(diamonds, aes(log10(carat), log10(price))) +  
 geom\_bin2d() +  
 facet\_wrap(~cut, nrow = 1)



mod\_coef <- coef(lm(log10(price) ~ log10(carat), data = diamonds))   
ggplot(diamonds, aes(log10(carat), log10(price))) +  
 geom\_bin2d() +  
 geom\_abline(intercept = mod\_coef[1], slope = mod\_coef[2],  
 colour = "white", size = 1) +  
 facet\_wrap(~cut, nrow = 1)

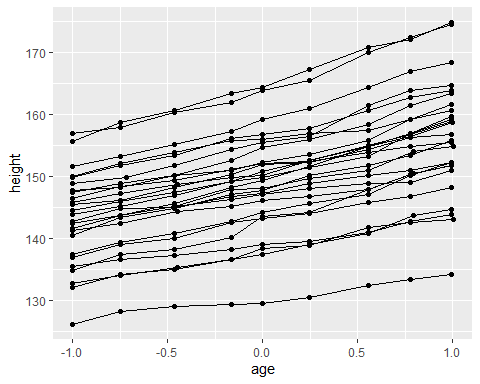
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



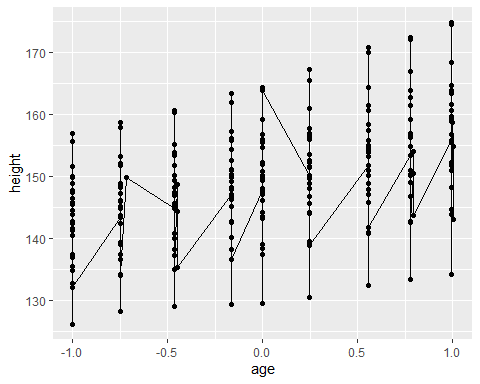
# 3.5  
data(Oxboys, package = "nlme")  
head(Oxboys)

## Grouped Data: height ~ age | Subject  
## Subject age height Occasion  
## 1 1 -1.0000 140.5 1  
## 2 1 -0.7479 143.4 2  
## 3 1 -0.4630 144.8 3  
## 4 1 -0.1643 147.1 4  
## 5 1 -0.0027 147.7 5  
## 6 1 0.2466 150.2 6

ggplot(Oxboys, aes(age, height, group = Subject)) +  
 geom\_point() +   
 geom\_line()

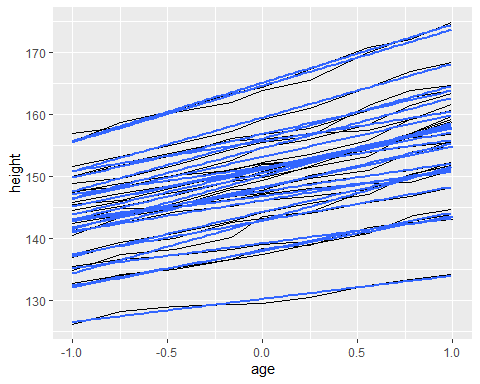


ggplot(Oxboys, aes(age, height)) +  
 geom\_point() +  
 geom\_line()



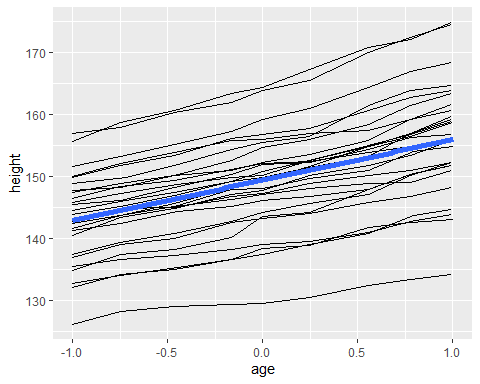
ggplot(Oxboys, aes(age, height, group = Subject)) +  
 geom\_line() +  
 geom\_smooth(method = "lm", se = FALSE)

## `geom\_smooth()` using formula = 'y ~ x'

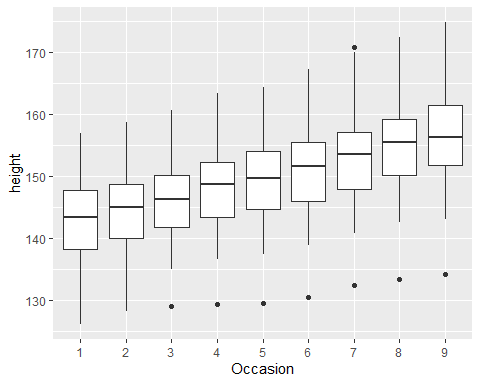


ggplot(Oxboys, aes(age, height)) +  
 geom\_line(aes(group = Subject)) +  
 geom\_smooth(method = "lm", size = 2, se = FALSE)

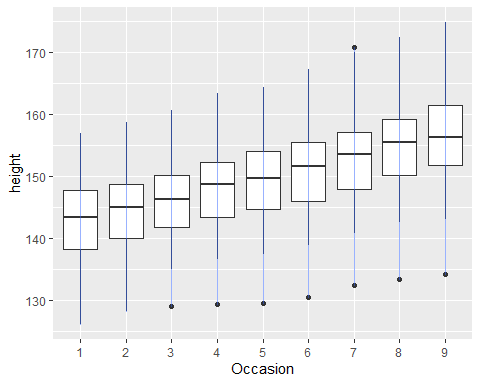
## `geom\_smooth()` using formula = 'y ~ x'



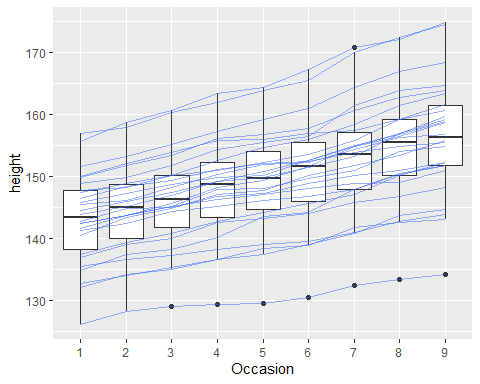
ggplot(Oxboys, aes(Occasion, height)) +  
 geom\_boxplot()



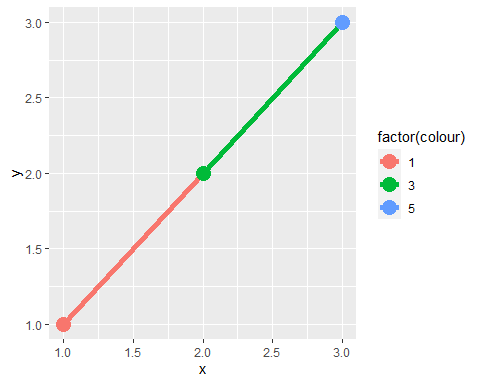
ggplot(Oxboys, aes(Occasion, height)) +   
 geom\_boxplot() +  
 geom\_line(colour = "#3366FF", alpha = 0.5)



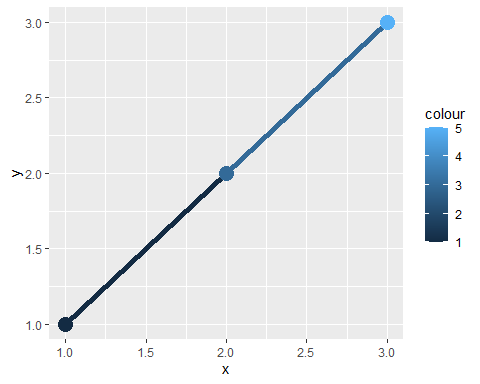
ggplot(Oxboys, aes(Occasion, height)) +  
 geom\_boxplot() +  
 geom\_line(aes(group = Subject), colour = "#3366FF", alpha = 0.5)



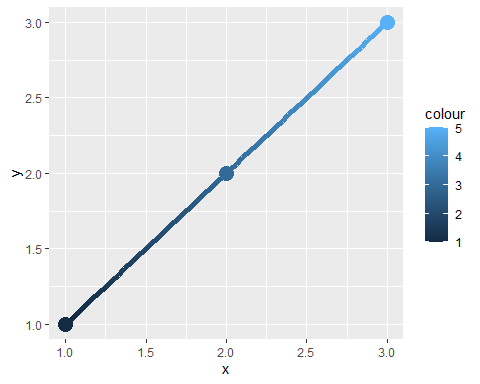
df <- data.frame(x = 1:3, y = 1:3, colour = c(1,3,5))  
ggplot(df, aes(x, y, colour = factor(colour))) +   
 geom\_line(aes(group = 1), size = 2) +  
 geom\_point(size = 5)



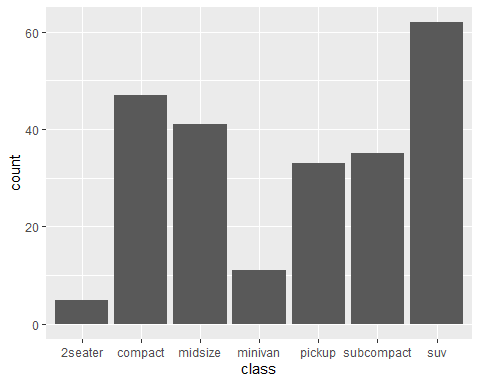
ggplot(df, aes(x, y, colour = colour)) +  
 geom\_line(aes(group = 1), size = 2) +   
 geom\_point(size = 5)



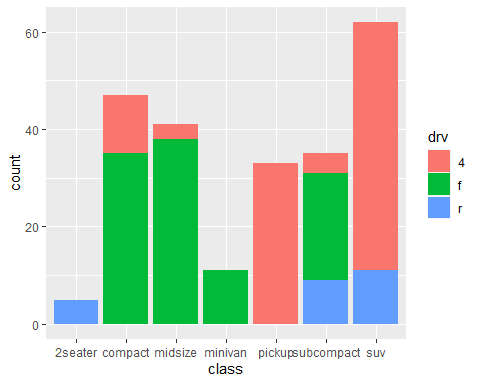
xgrid <- with(df, seq(min(x), max(x), length = 50))  
interp <- data.frame(  
 x = xgrid,  
 y = approx(df$x, df$y, xout = xgrid)$y,  
 colour = approx(df$x, df$colour, xout = xgrid)$y  
)  
ggplot(interp, aes(x, y, colour = colour)) +  
 geom\_line(size = 2) +  
 geom\_point(data = df, size = 5)



ggplot(mpg, aes(class)) +  
 geom\_bar()

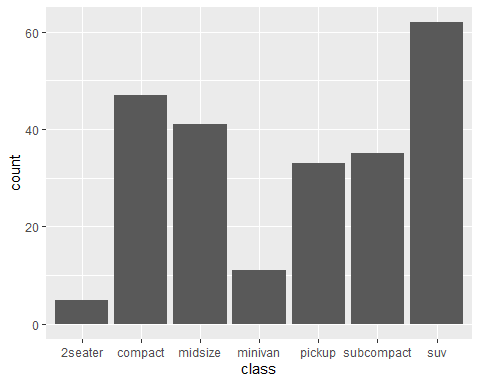


ggplot(mpg, aes(class, fill = drv)) +   
 geom\_bar()

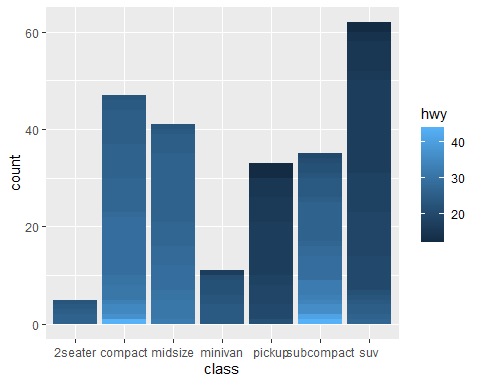


ggplot(mpg, aes(class, fill = hwy)) +  
 geom\_bar()

## Warning: The following aesthetics were dropped during statistical transformation: fill  
## ℹ This can happen when ggplot fails to infer the correct grouping structure in  
## the data.  
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical  
## variable into a factor?

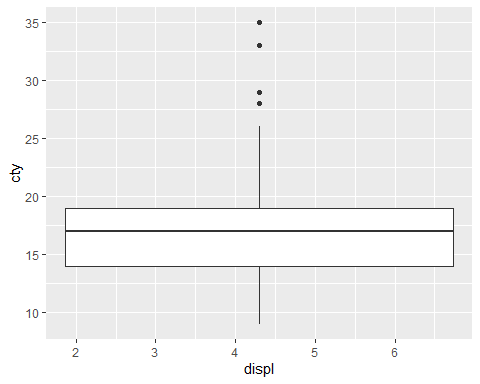


ggplot(mpg, aes(class, fill = hwy, group = hwy)) +  
 geom\_bar()

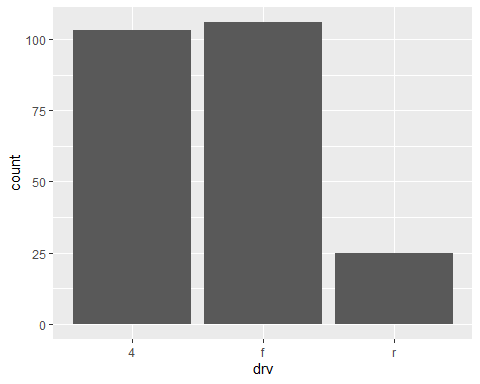


ggplot(mpg, aes(displ, cty)) +  
 geom\_boxplot()

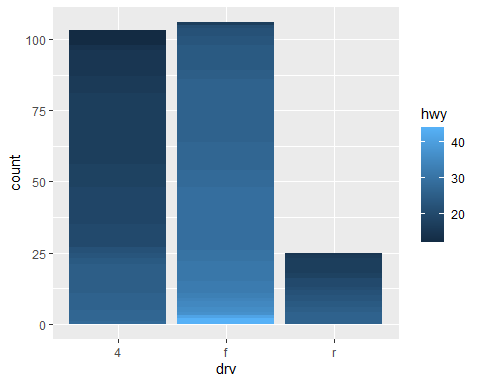
## Warning: Continuous x aesthetic  
## ℹ did you forget `aes(group = ...)`?



ggplot(mpg, aes(drv)) +  
 geom\_bar()



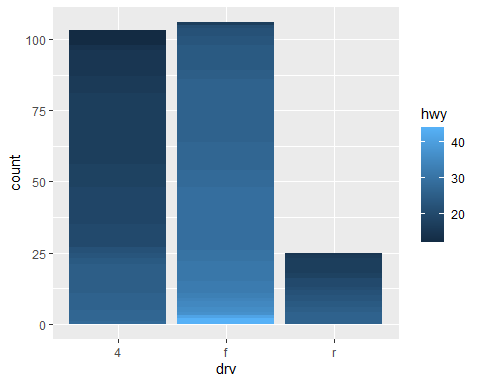
ggplot(mpg, aes(drv, fill = hwy, group = hwy)) +  
 geom\_bar()



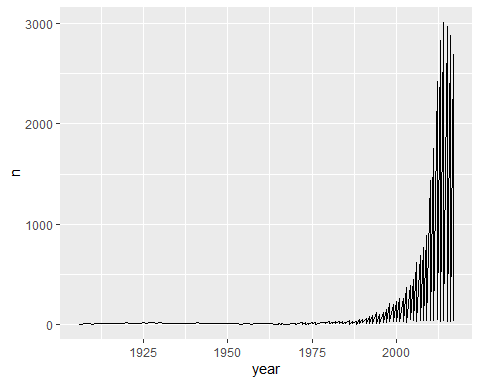
library(dplyr)

##   
## Attaching package: 'dplyr'  
##   
## The following object is masked from 'package:nlme':  
##   
## collapse  
##   
## The following objects are masked from 'package:stats':  
##   
## filter, lag  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

mpg2 <- mpg %>% arrange(hwy) %>% mutate(id = seq\_along(hwy))  
ggplot(mpg2, aes(drv, fill = hwy, group = id)) +  
 geom\_bar()

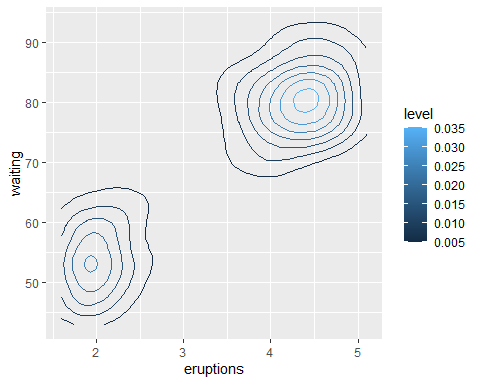


library(babynames)  
hadley <- dplyr::filter(babynames, name == "Hadley")  
ggplot(hadley, aes(year, n)) +  
 geom\_line()

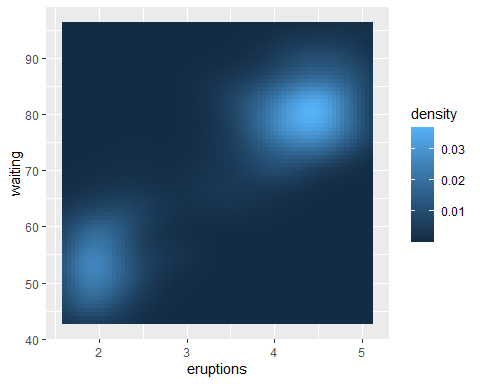


# 3.6  
ggplot(faithfuld, aes(eruptions, waiting)) +   
 geom\_contour(aes(z = density, colour = ..level..))

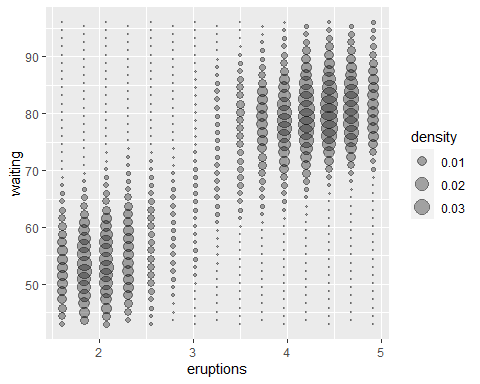
## Warning: The dot-dot notation (`..level..`) was deprecated in ggplot2 3.4.0.  
## ℹ Please use `after\_stat(level)` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



ggplot(faithfuld, aes(eruptions, waiting)) +  
 geom\_raster(aes(fill = density))



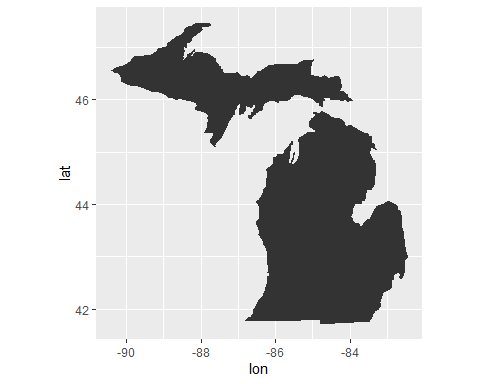
small <- faithfuld[seq(1, nrow(faithfuld), by = 10), ]  
ggplot(small, aes(eruptions, waiting)) +  
 geom\_point(aes(size = density), alpha = 1/3) +  
 scale\_size\_area()



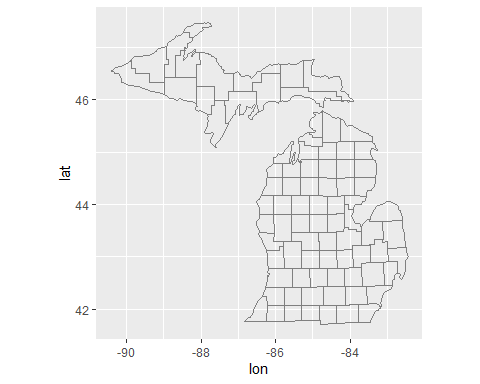
# 3.7  
# Taiwan\_map <- GetMap(center = c(lat = 22.996758, lon = 120.216857), zoom = 18, destfile = "Taiwan.png")  
# NCKU.lat <- c()  
mi\_counties <- map\_data("county", "michigan") %>%  
 select(lon = long, lat, group, id = subregion)  
head(mi\_counties)

## lon lat group id  
## 1 -83.88675 44.85686 1 alcona  
## 2 -83.36536 44.86832 1 alcona  
## 3 -83.36536 44.86832 1 alcona  
## 4 -83.33098 44.83968 1 alcona  
## 5 -83.30806 44.80530 1 alcona  
## 6 -83.30233 44.77665 1 alcona

ggplot(mi\_counties, aes(lon, lat)) +  
 geom\_polygon(aes(group = group)) +  
 coord\_quickmap()



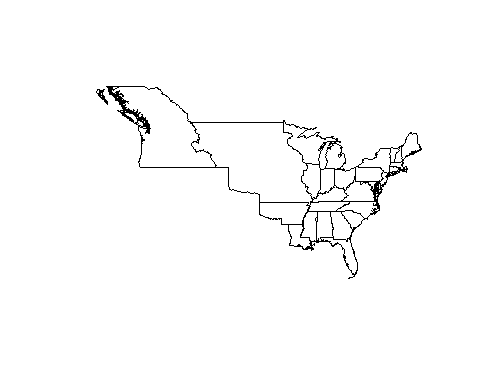
ggplot(mi\_counties, aes(lon, lat)) +  
 geom\_polygon(aes(group = group), fill = NA, colour = "grey50") +  
 coord\_quickmap()



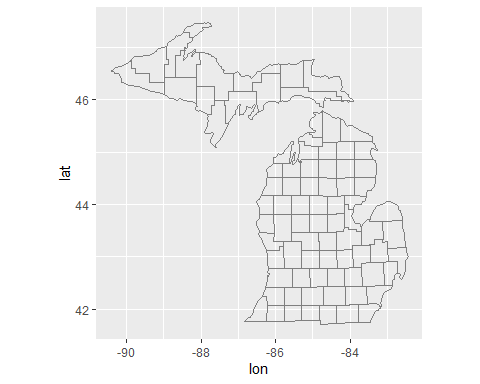
library(USAboundaries)  
library(sf)

## Linking to GEOS 3.11.2, GDAL 3.7.2, PROJ 9.3.0; sf\_use\_s2() is TRUE

c18 <- us\_states(as.Date("1820-01-01"))  
plot(st\_geometry(c18))



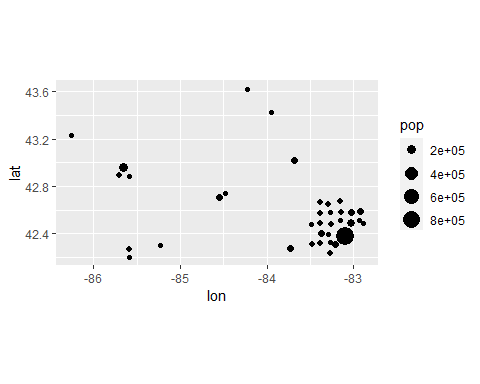
# c18df <- fortify(c18)  
  
ggplot(mi\_counties, aes(lon, lat)) +  
 geom\_polygon(aes(group = group), colour = "grey50", fill = NA) +  
 coord\_quickmap()



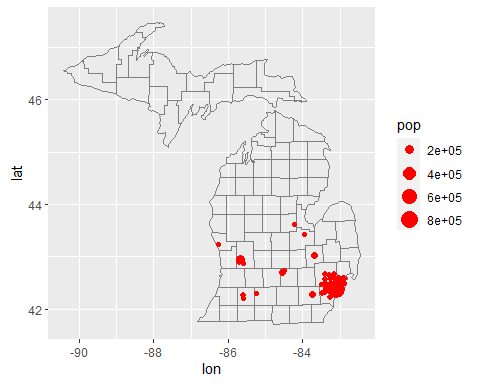
mi\_cities <- maps::us.cities %>%  
 tibble::as\_tibble() %>%  
 filter(country.etc =="MI") %>%  
 select(-country.etc, lon = long) %>%  
 arrange(desc(pop))  
mi\_cities

## # A tibble: 36 × 5  
## name pop lat lon capital  
## <chr> <int> <dbl> <dbl> <int>  
## 1 Detroit MI 871789 42.4 -83.1 0  
## 2 Grand Rapids MI 193006 43.0 -85.7 0  
## 3 Warren MI 132537 42.5 -83.0 0  
## 4 Sterling Heights MI 127027 42.6 -83.0 0  
## 5 Lansing MI 117236 42.7 -84.6 2  
## 6 Flint MI 115691 43.0 -83.7 0  
## 7 Ann Arbor MI 113716 42.3 -83.7 0  
## 8 Clinton MI 100517 42.6 -82.9 0  
## 9 Livonia MI 97722 42.4 -83.4 0  
## 10 Dearborn MI 94681 42.3 -83.2 0  
## # ℹ 26 more rows

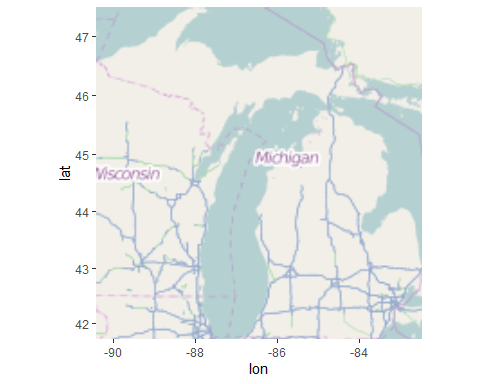
ggplot(mi\_cities, aes(lon, lat)) +  
 geom\_point(aes(size = pop)) +  
 scale\_size\_area() +  
 coord\_quickmap()



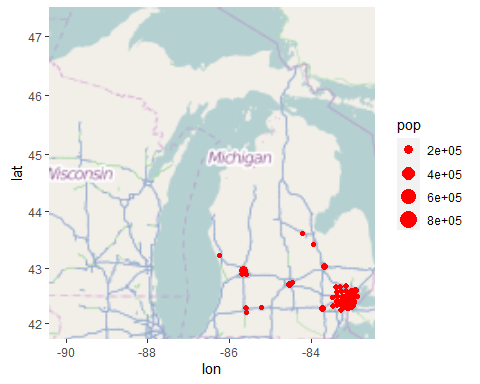
ggplot(mi\_cities, aes(lon, lat)) +  
 geom\_polygon(aes(group = group), mi\_counties, fill = NA, colour = "grey50") +  
 geom\_point(aes(size = pop), colour = "red") +  
 scale\_size\_area() +  
 coord\_quickmap()



# if (file.exists("mi\_raster.rds")) {  
# mi\_raster <- readRDS("mi\_raster.rds")  
# } else {  
# bbox <- c(  
# min(mi\_counties$lon), min(mi\_counties$lat),  
# max(mi\_counties$lon), max(mi\_counties$lat)  
# )  
# mi\_raster <- ggmap::get\_openstreetmap(bbox, scale = 8735660)  
# saveRDS(mi\_raster, "mi\_raster.rds")  
# }  
  
mi\_raster <- readRDS("C:/Users/wumin/Downloads/mi\_raster.rds")  
  
ggmap::ggmap(mi\_raster)



ggmap::ggmap(mi\_raster) +  
 geom\_point(aes(size = pop), mi\_cities, colour = "red") +  
 scale\_size\_area()



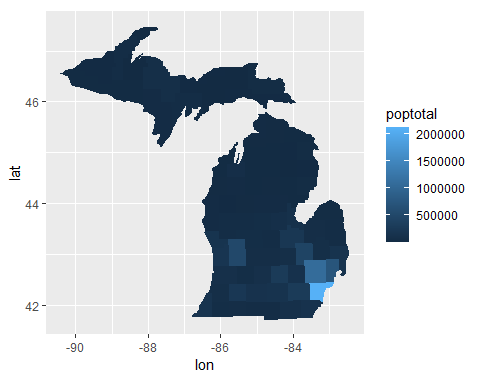
# df <- as.data.frame(raster::rasterToPoints(x))  
# names(df) <- c("lon", "lat", "x")  
#   
# ggplot(df, aes(lon, lat)) +  
# geom\_raster(aes(fill = x))  
  
mi\_census <- midwest %>%  
 tibble::as\_tibble() %>%  
 filter(state =="MI") %>%  
 mutate(county = tolower(county)) %>%  
 select(county, area, poptotal, percwhite, percblack)  
mi\_census

## # A tibble: 83 × 5  
## county area poptotal percwhite percblack  
## <chr> <dbl> <int> <dbl> <dbl>  
## 1 alcona 0.041 10145 98.8 0.266   
## 2 alger 0.051 8972 93.9 2.37   
## 3 allegan 0.049 90509 95.9 1.60   
## 4 alpena 0.034 30605 99.2 0.114   
## 5 antrim 0.031 18185 98.4 0.126   
## 6 arenac 0.021 14931 98.4 0.0670  
## 7 baraga 0.054 7954 87.6 0.616   
## 8 barry 0.034 50057 98.7 0.208   
## 9 bay 0.026 111723 96.4 1.11   
## 10 benzie 0.02 12200 97.2 0.246   
## # ℹ 73 more rows

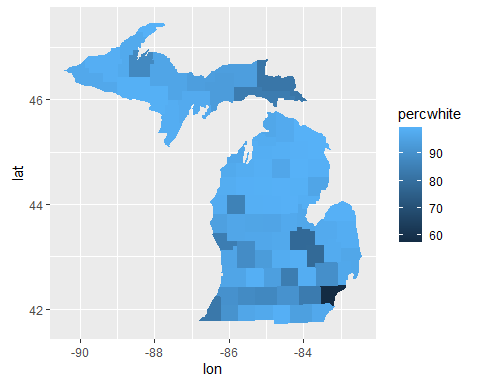
census\_counties <- left\_join(mi\_census, mi\_counties, by = c("county" ="id"))  
census\_counties

## # A tibble: 1,472 × 8  
## county area poptotal percwhite percblack lon lat group  
## <chr> <dbl> <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 alcona 0.041 10145 98.8 0.266 -83.9 44.9 1  
## 2 alcona 0.041 10145 98.8 0.266 -83.4 44.9 1  
## 3 alcona 0.041 10145 98.8 0.266 -83.4 44.9 1  
## 4 alcona 0.041 10145 98.8 0.266 -83.3 44.8 1  
## 5 alcona 0.041 10145 98.8 0.266 -83.3 44.8 1  
## 6 alcona 0.041 10145 98.8 0.266 -83.3 44.8 1  
## 7 alcona 0.041 10145 98.8 0.266 -83.3 44.7 1  
## 8 alcona 0.041 10145 98.8 0.266 -83.3 44.7 1  
## 9 alcona 0.041 10145 98.8 0.266 -83.3 44.7 1  
## 10 alcona 0.041 10145 98.8 0.266 -83.3 44.6 1  
## # ℹ 1,462 more rows

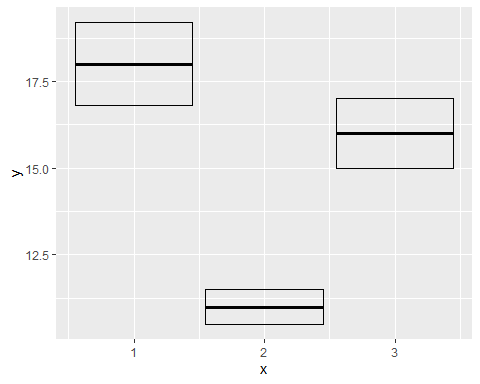
ggplot(census\_counties, aes(lon, lat, group = county)) +  
 geom\_polygon(aes(fill = poptotal)) +  
 coord\_quickmap()



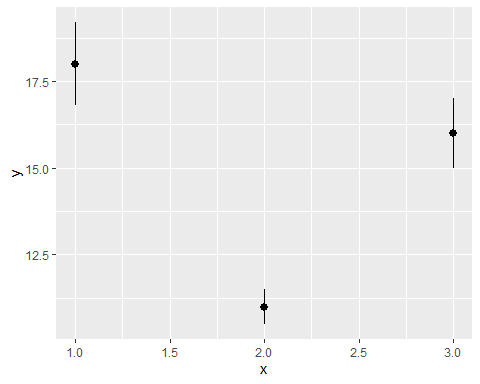
ggplot(census\_counties, aes(lon, lat, group = county)) +  
 geom\_polygon(aes(fill = percwhite)) +  
 coord\_quickmap()



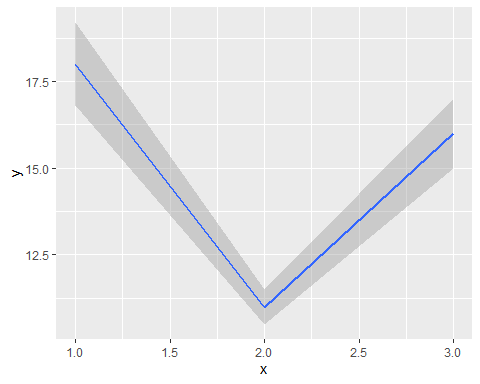
#3.8  
y <- c(18, 11, 16)   
df <- data.frame(x = 1:3, y = y, se = c(1.2, 0.5, 1.0))  
base <- ggplot(df, aes(x, y, ymin = y - se, ymax = y + se))  
  
base + geom\_crossbar()



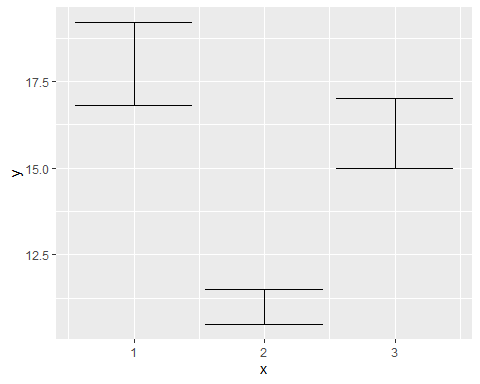
base + geom\_pointrange()



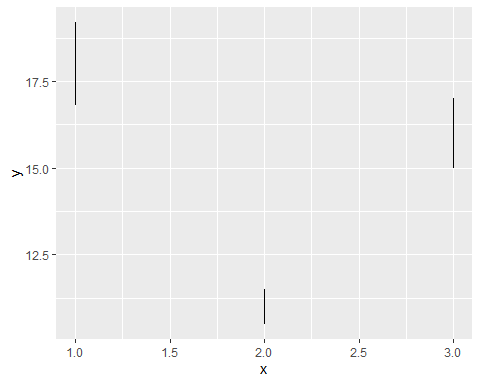
base + geom\_smooth(stat = "identity")



base + geom\_errorbar()



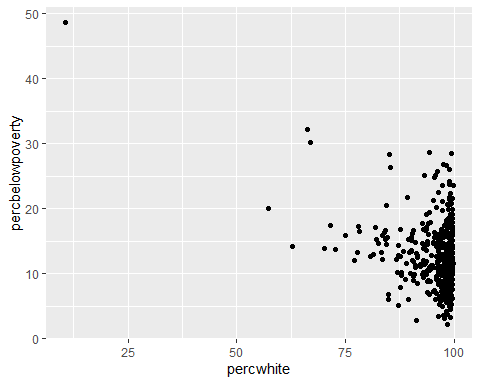
base + geom\_linerange()



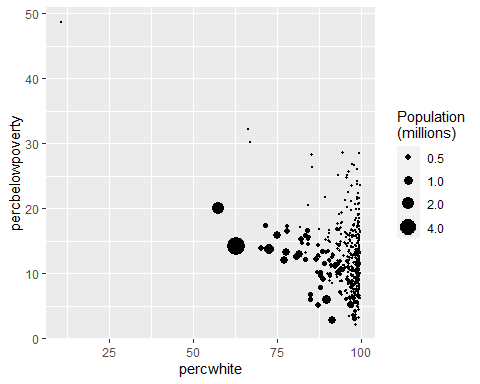
base + geom\_ribbon()



#3.9  
ggplot(midwest, aes(percwhite, percbelowpoverty)) +  
 geom\_point()

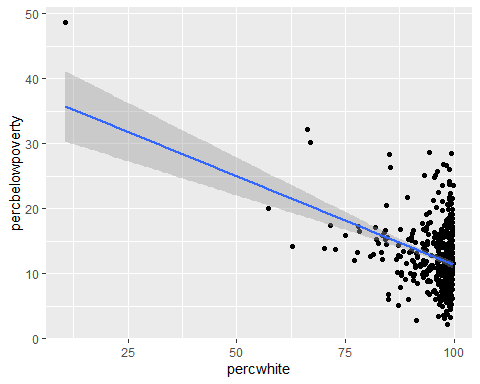


ggplot(midwest, aes(percwhite, percbelowpoverty)) +  
 geom\_point(aes(size = poptotal / 1e6)) +  
 scale\_size\_area("Population\n(millions)", breaks = c(0.5, 1, 2, 4))



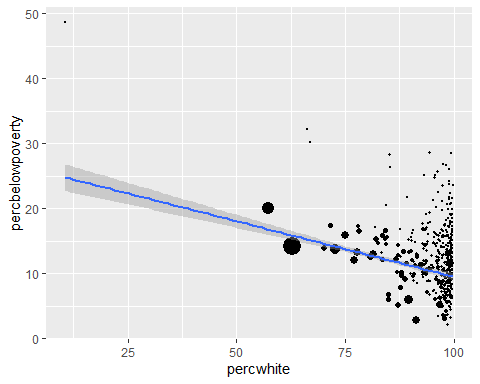
ggplot(midwest, aes(percwhite, percbelowpoverty)) +  
 geom\_point() +  
 geom\_smooth(method = lm, size = 1)

## `geom\_smooth()` using formula = 'y ~ x'

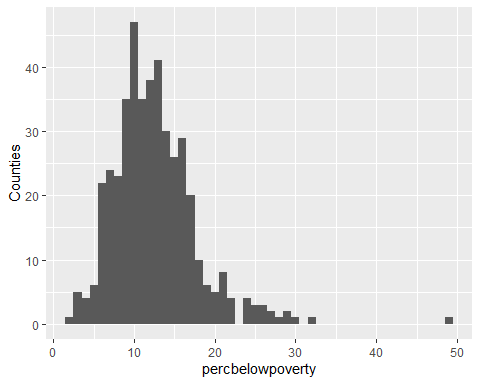


ggplot(midwest, aes(percwhite, percbelowpoverty)) +  
 geom\_point(aes(size = poptotal / 1e6)) +  
 geom\_smooth(aes(weight = poptotal), method = lm, size = 1) +  
 scale\_size\_area(guide = "none")

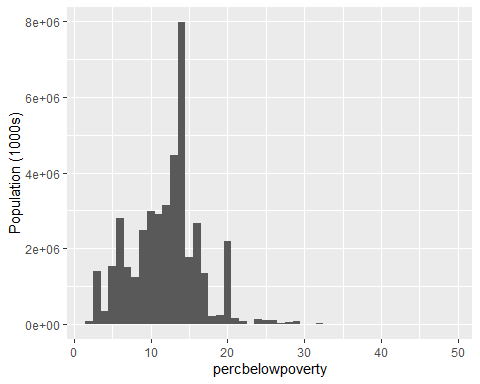
## `geom\_smooth()` using formula = 'y ~ x'



ggplot(midwest, aes(percbelowpoverty)) +  
 geom\_histogram(binwidth = 1) +  
 ylab("Counties")



ggplot(midwest, aes(percbelowpoverty)) +  
 geom\_histogram(aes(weight = poptotal), binwidth = 1) +  
 ylab("Population (1000s)")

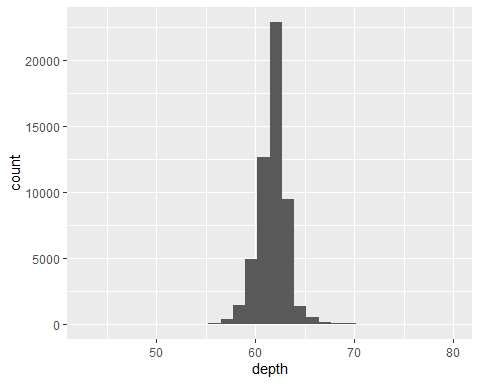


#3.10  
diamonds

## # A tibble: 53,940 × 10  
## carat cut color clarity depth table price x y z  
## <dbl> <ord> <ord> <ord> <dbl> <dbl> <int> <dbl> <dbl> <dbl>  
## 1 0.23 Ideal E SI2 61.5 55 326 3.95 3.98 2.43  
## 2 0.21 Premium E SI1 59.8 61 326 3.89 3.84 2.31  
## 3 0.23 Good E VS1 56.9 65 327 4.05 4.07 2.31  
## 4 0.29 Premium I VS2 62.4 58 334 4.2 4.23 2.63  
## 5 0.31 Good J SI2 63.3 58 335 4.34 4.35 2.75  
## 6 0.24 Very Good J VVS2 62.8 57 336 3.94 3.96 2.48  
## 7 0.24 Very Good I VVS1 62.3 57 336 3.95 3.98 2.47  
## 8 0.26 Very Good H SI1 61.9 55 337 4.07 4.11 2.53  
## 9 0.22 Fair E VS2 65.1 61 337 3.87 3.78 2.49  
## 10 0.23 Very Good H VS1 59.4 61 338 4 4.05 2.39  
## # ℹ 53,930 more rows

ggplot(diamonds, aes(depth)) +  
 geom\_histogram()

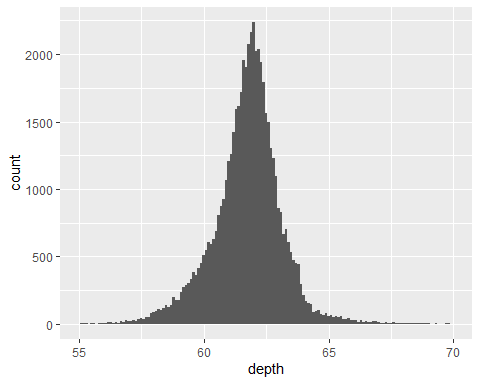
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



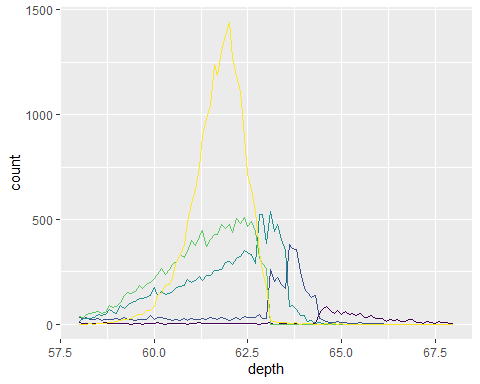
ggplot(diamonds, aes(depth)) +  
 geom\_histogram(binwidth = 0.1) +  
 xlim(55, 70)

## Warning: Removed 45 rows containing non-finite values (`stat\_bin()`).

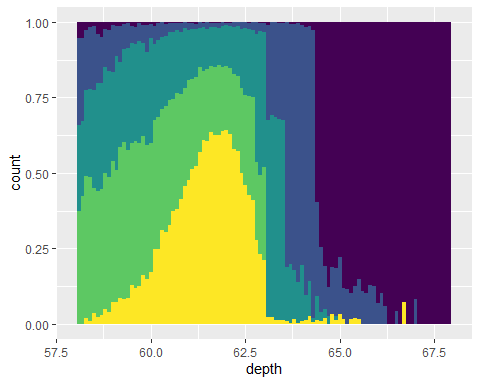
## Warning: Removed 2 rows containing missing values (`geom\_bar()`).



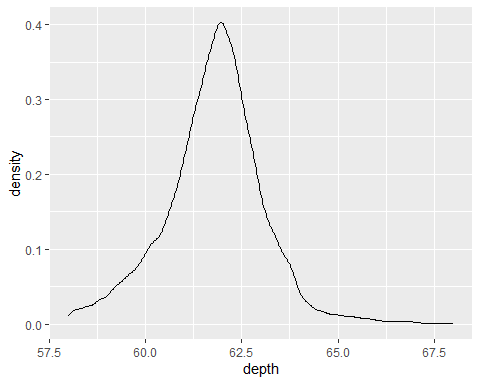
ggplot(diamonds, aes(depth)) +  
 geom\_freqpoly(aes(colour = cut), binwidth = 0.1, na.rm = TRUE) +  
 xlim(58, 68) +  
 theme(legend.position = "none")



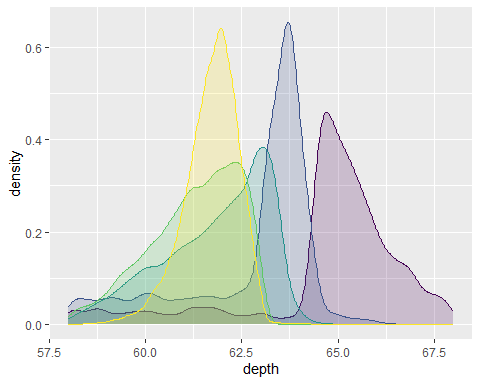
ggplot(diamonds, aes(depth)) +  
 geom\_histogram(aes(fill = cut), binwidth = 0.1, position = "fill",   
 na.rm = TRUE) +  
 xlim(58, 68) +  
 theme(legend.position = "none")



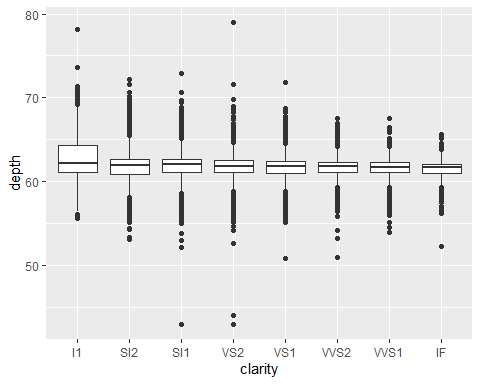
ggplot(diamonds, aes(depth)) +  
 geom\_density(na.rm = TRUE) +  
 xlim(58, 68) +  
 theme(legend.position = "none")



ggplot(diamonds, aes(depth, fill = cut, colour = cut)) +  
 geom\_density(alpha = 0.2, na.rm = TRUE) +  
 xlim(58, 68) +  
 theme(legend.position = "none")

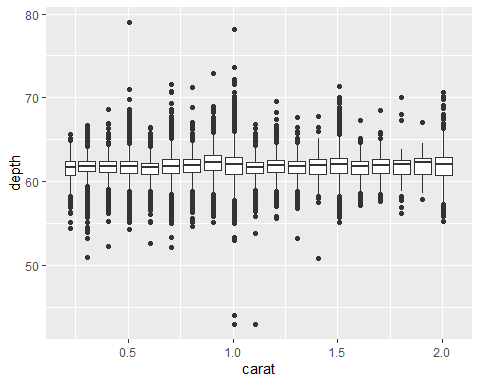


ggplot(diamonds, aes(clarity, depth)) +  
 geom\_boxplot()

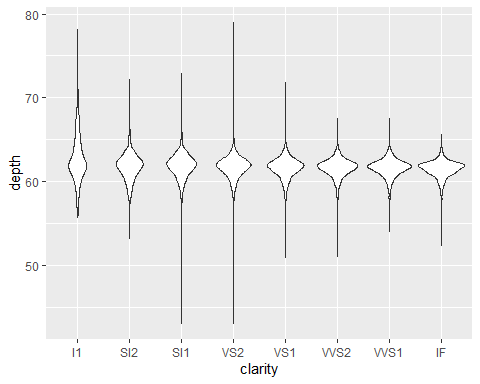


ggplot(diamonds, aes(carat, depth)) +  
 geom\_boxplot(aes(group = cut\_width(carat, 0.1))) +  
 xlim(NA, 2.05)

## Warning: Removed 997 rows containing missing values (`stat\_boxplot()`).



ggplot(diamonds, aes(clarity, depth)) +  
 geom\_violin()

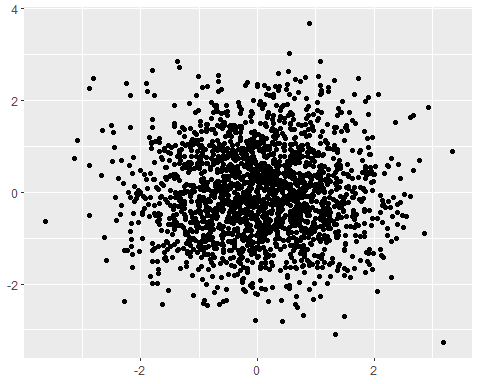


ggplot(diamonds, aes(carat, depth)) +  
 geom\_violin(aes(group = cut\_width(carat, 0.1))) +  
 xlim(NA, 2.05)

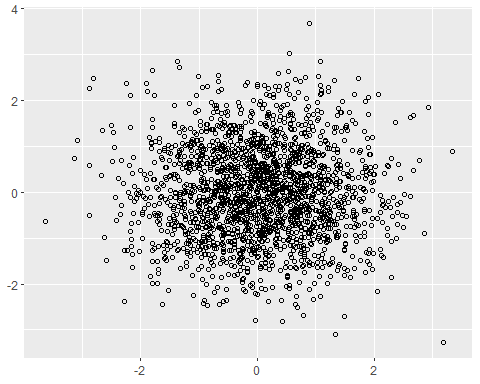
## Warning: Removed 997 rows containing non-finite values (`stat\_ydensity()`).



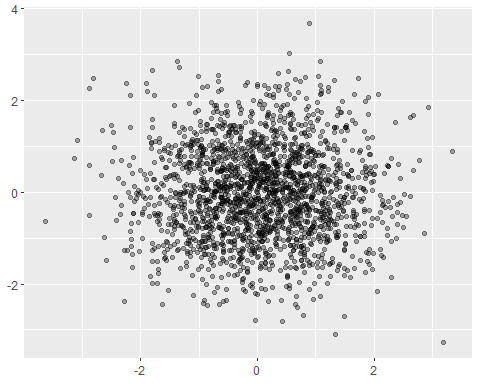
# 3.12  
df <- data.frame(x = rnorm(2000), y = rnorm(2000))  
norm <- ggplot(df, aes(x, y)) + xlab(NULL) + ylab(NULL)  
  
norm + geom\_point()



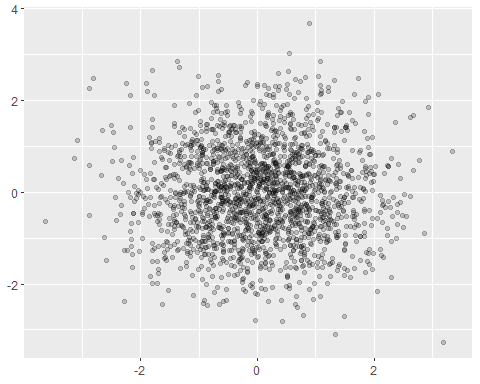
norm + geom\_point(shape = 1)



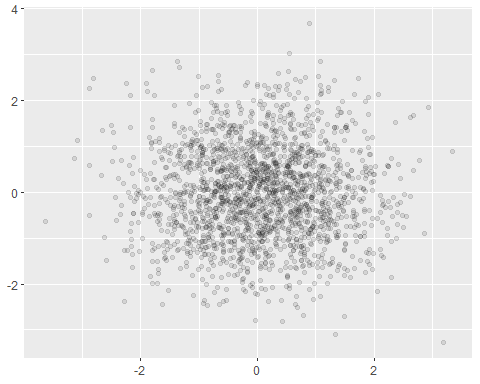
norm + geom\_point(alpha = 1 / 3)



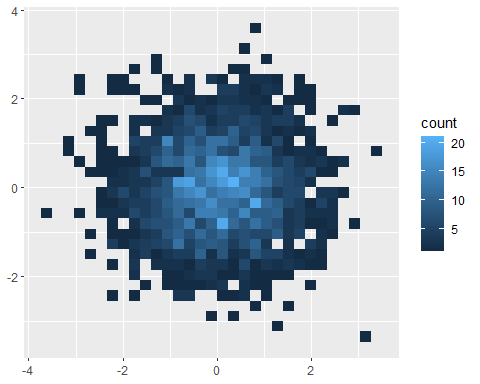
norm + geom\_point(alpha = 1 / 5)



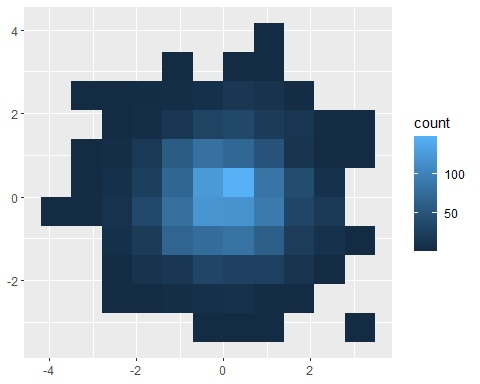
norm + geom\_point(alpha = 1 / 10)



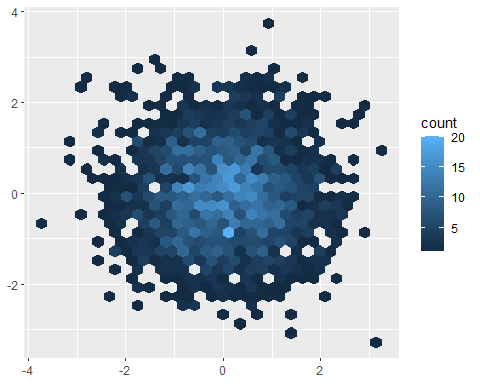
norm + geom\_bin2d()



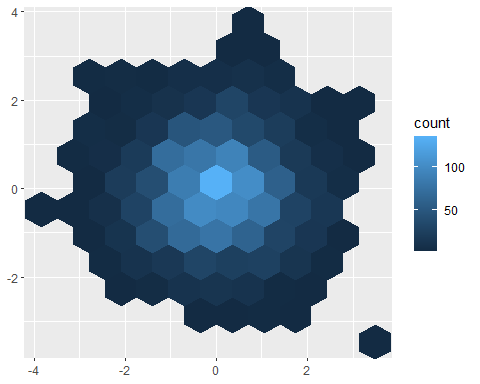
norm + geom\_bin2d(bins = 10)



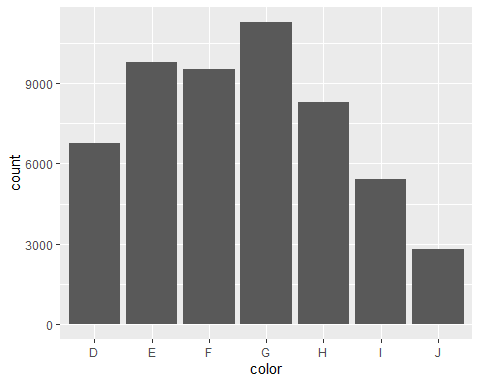
norm + geom\_hex()



norm + geom\_hex(bins = 10)



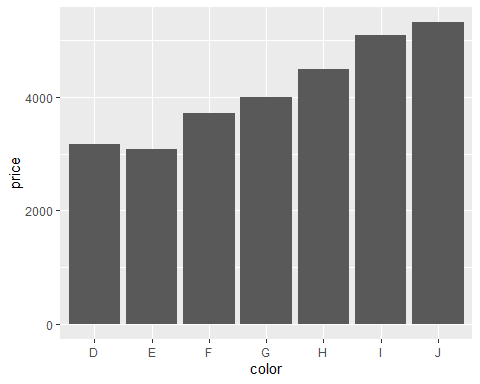
ggplot(diamonds, aes(color)) +  
 geom\_bar()



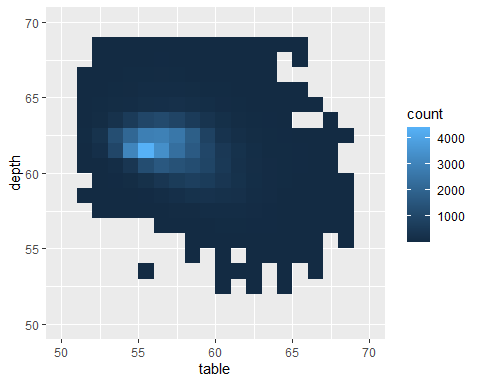
ggplot(diamonds, aes(color, price)) +  
 geom\_bar(stat = "summary\_bin", fun.y = mean)

## Warning in geom\_bar(stat = "summary\_bin", fun.y = mean): Ignoring unknown  
## parameters: `fun.y`

## No summary function supplied, defaulting to `mean\_se()`  
## No summary function supplied, defaulting to `mean\_se()`  
## No summary function supplied, defaulting to `mean\_se()`  
## No summary function supplied, defaulting to `mean\_se()`  
## No summary function supplied, defaulting to `mean\_se()`  
## No summary function supplied, defaulting to `mean\_se()`  
## No summary function supplied, defaulting to `mean\_se()`



ggplot(diamonds, aes(table, depth)) +  
 geom\_bin2d(binwidth = 1, na.rm = TRUE) +  
 xlim(50, 70) +  
 ylim(50, 70)



ggplot(diamonds, aes(table, depth, z = price)) +  
 geom\_raster(binwidth = 1, stat = "summary\_2d", fun = mean,  
 na.rm = TRUE) +   
 xlim(50, 70) +  
 ylim(50, 70)

## Warning: Raster pixels are placed at uneven horizontal intervals and will be shifted  
## ℹ Consider using `geom\_tile()` instead.

## Warning: Raster pixels are placed at uneven horizontal intervals and will be shifted  
## ℹ Consider using `geom\_tile()` instead.

