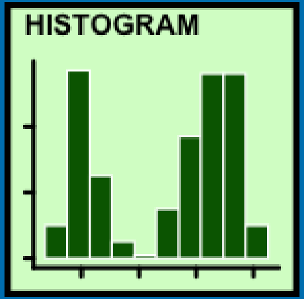
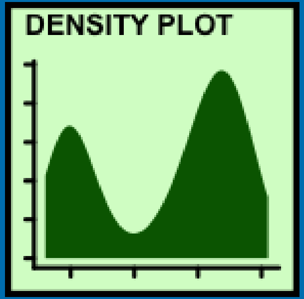
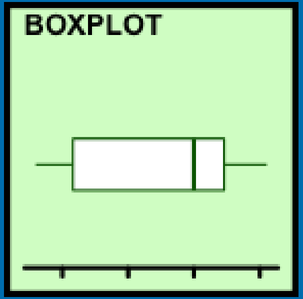
Data Visualisation with R

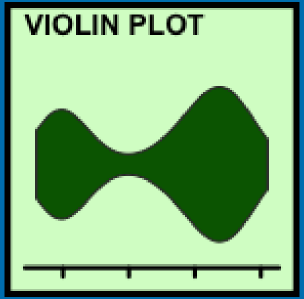
Workshop Part 1

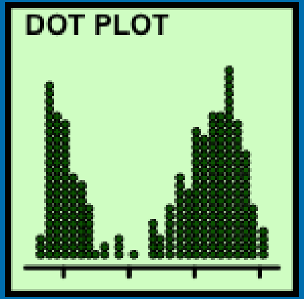
Catalogue

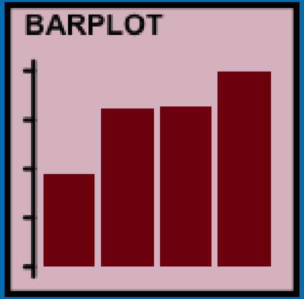

HISTOGRAM

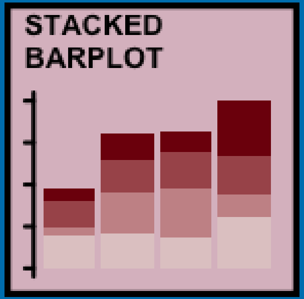

DENSITY PLOT

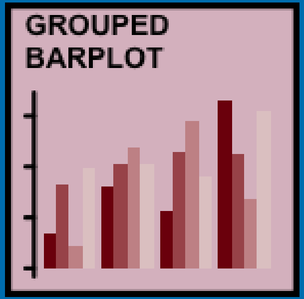

BOXPLOT

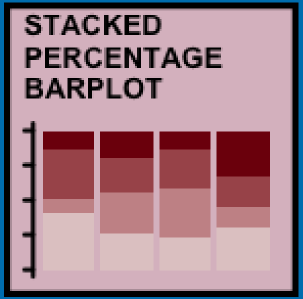

VIOLIN PLOT

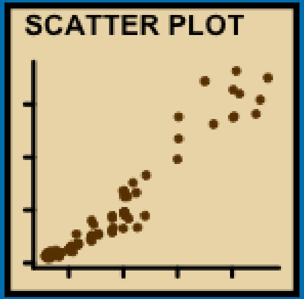

DOT PLOT


BARPLOT

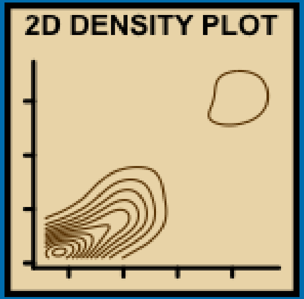

STACKED BARPLOT


GROUPED BARPLOT

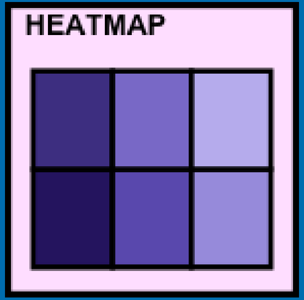

STACKED PERCENTAGE BARPLOT

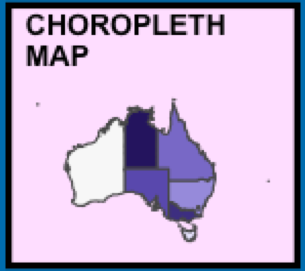

SCATTER PLOT


HEX PLOT


2D DENSITY PLOT


LINE PLOT


HEATMAP


CHOROPLETH MAP

Illustrative data 💰 Wages

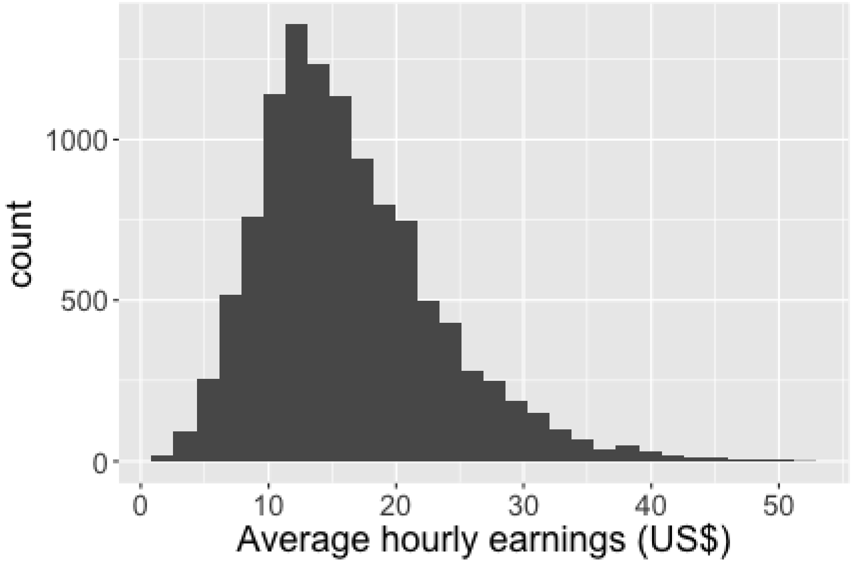
- Let's look at a survey of average hourly earnings in United States where each observation is an individual.

```r
data(CPSch3, package = "Ecdat")
str(CPSch3)
```

```
## 'data.frame':    11130 obs. of  3 variables:
##  $ year: int  1992 1992 1992 1992 1992 1992 1992 1992 199
##  $ ahe : num  13 11.6 17.4 10.1 16.8 ...
##  $ sex : Factor w/ 2 levels "male","female": 1 1 1 2 1 2
```

A histogram with geom_histogram( )

```
ggplot(data = CPSch3,
       mapping = aes(x = ahe)) +
  geom_histogram() +
  labs(x = "Average hourly earnings (US$)")
```



▸ Back to catalogue

Layer data

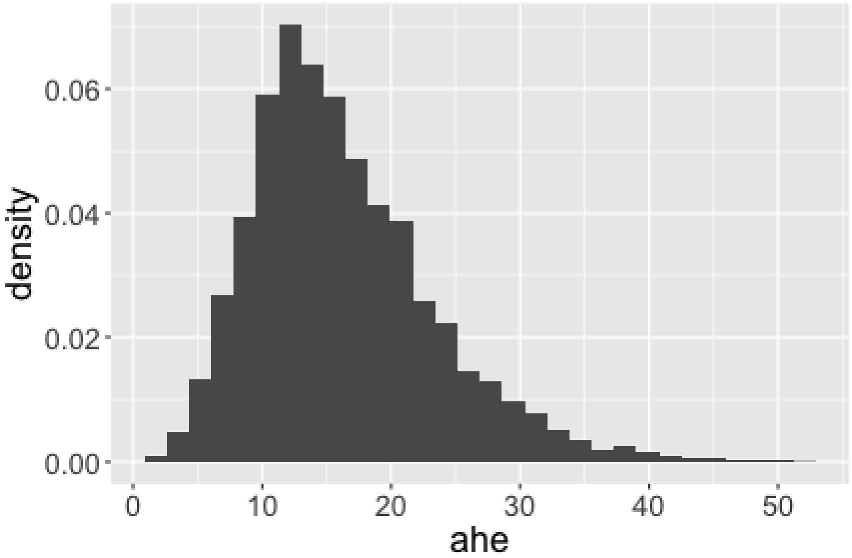
```r
g <- ggplot(data = CPSch3,
       mapping = aes(x = ahe)) +
  geom_histogram()

layer_data(g, 1)
```

```
##       y count         x       xmin      xmax      density      ncount    ndensity flipped_aes
## 1    19    19  1.734720  0.8673601  2.602080 0.0009840768 0.014011799 0.014011799       FALSE
## 2    91    91  3.469440  2.6020802  4.336800 0.0047132101 0.067109145 0.067109145       FALSE
## 3   255   255  5.204160  4.3368004  6.071521 0.0132073470 0.188053097 0.188053097       FALSE
## 4   514   514  6.938881  6.0715206  7.806241 0.0266218681 0.379056047 0.379056047       FALSE
## 5   760   760  8.673601  7.8062407  9.540961 0.0393630735 0.560471976 0.560471976       FALSE
## 6  1143  1143 10.408321  9.5409609 11.275681 0.0591999908 0.842920354 0.842920354       FALSE
## 7  1356  1356 12.143041 11.2756811 13.010401 0.0702320101 1.000000000 1.000000000       FALSE
## 8  1232  1232 13.877761 13.0104012 14.745121 0.0638096139 0.908554572 0.908554572       FALSE
## 9  1137  1137 15.612481 14.7451214 16.479842 0.0588892297 0.838495575 0.8384955
## 10  941   941 17.347202 16.4798416 18.214562 0.0487377002 0.693952802 0.6939528
```

▸ Back to catalogue

Layer data

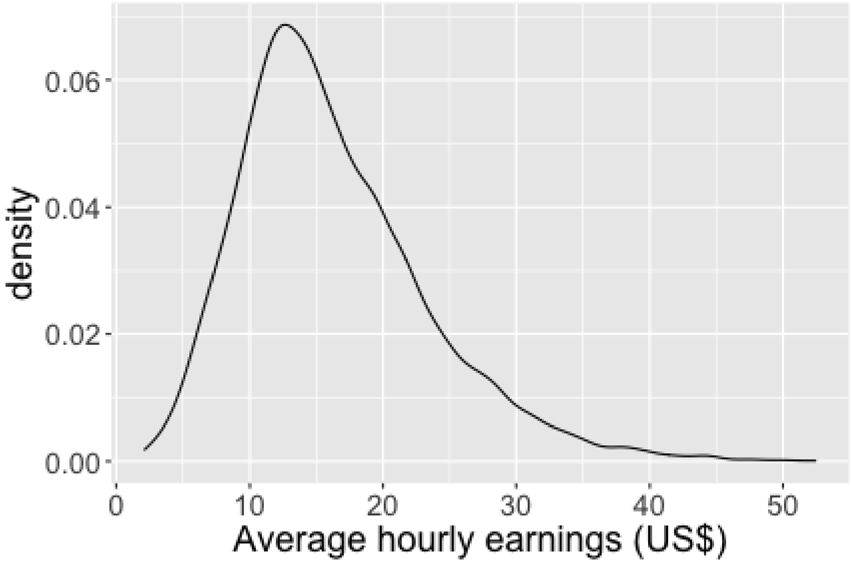
```
ggplot(data = CPSch3,
       mapping = aes(x = ahe)) +
  geom_histogram(aes(y = after_stat(density)))
```



▸ Back to catalogue

A density plot with geom_density()

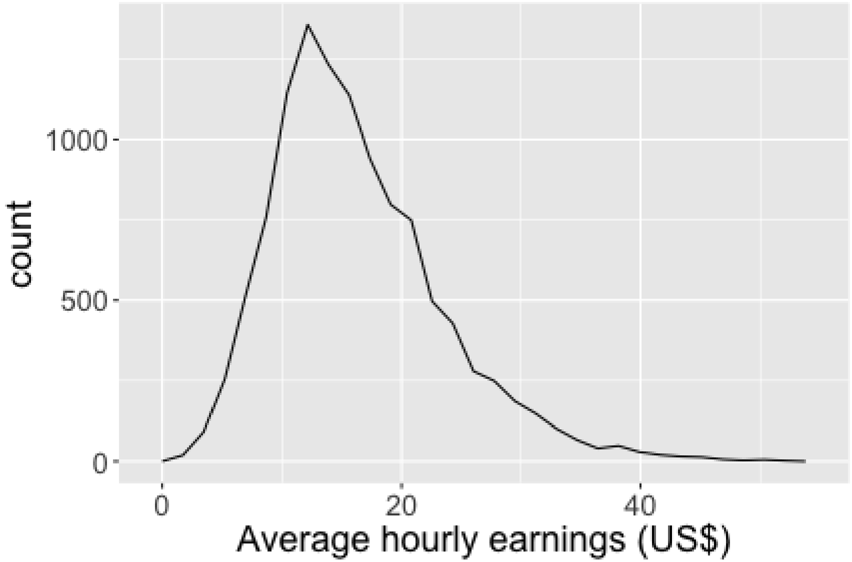
```
ggplot(data = CPSch3,
       mapping = aes(x = ahe)) +
  geom_density() +
  labs(x = "Average hourly earnings (US$)")
```



▸ Back to catalogue

A frequency polygon with geom_freqpoly()

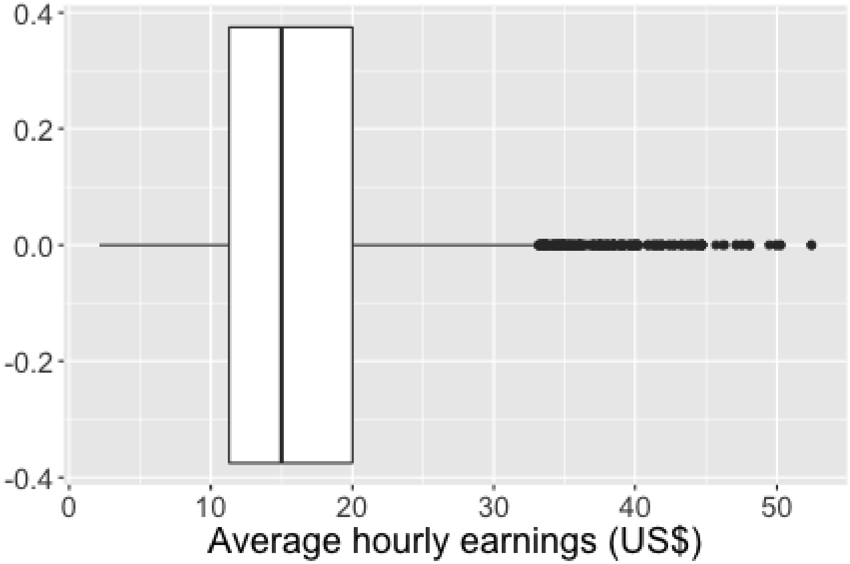
```
ggplot(data = CPSch3,
       mapping = aes(x = ahe)) +
  geom_freqpoly() +
  labs(x = "Average hourly earnings (US$)")
```



Back to catalogue

A boxplot with geom_boxplot()

```
ggplot(data = CPSch3,
       mapping = aes(x = ahe)) +
  geom_boxplot() +
  labs(x = "Average hourly earnings (US$)")
```



Back to catalogue

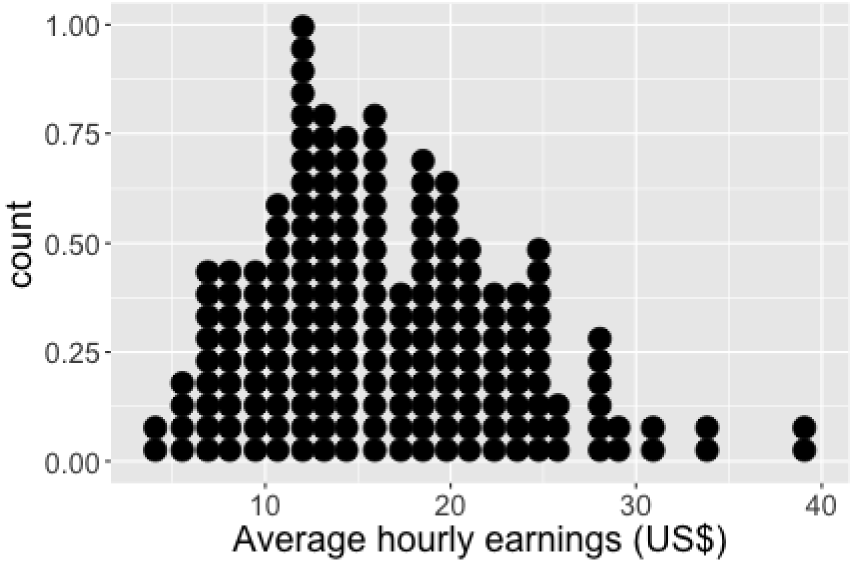
A violin plot with geom_violin()

```
ggplot(data = CPSch3,
       mapping = aes(x = ahe, y = "")) +
  geom_violin() +
  labs(x = "Average hourly earnings (US$)")
```

▸ Back to catalogue

A dotplot with geom_dotplot()

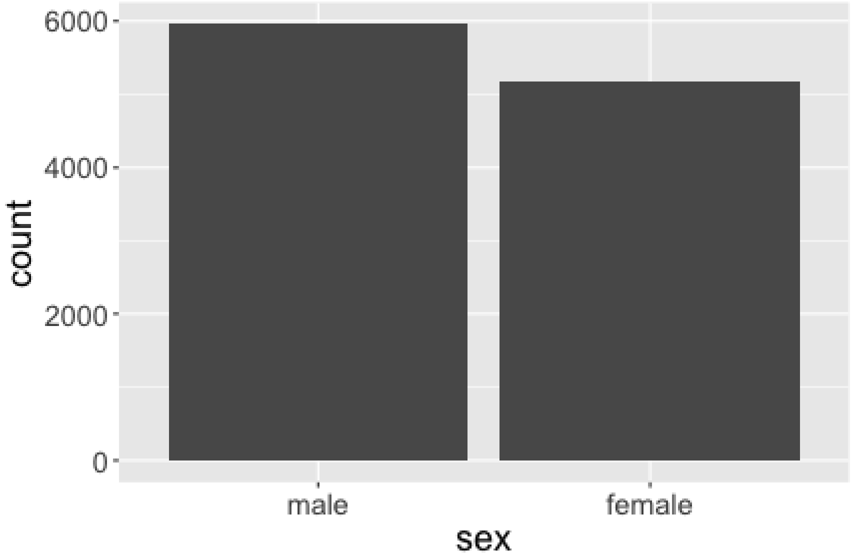
```
ggplot(data = dplyr::sample_n(CPSch3, 200),
       mapping = aes(x = ahe)) +
  geom_dotplot() +
  labs(x = "Average hourly earnings (US$)")
```



Back to catalogue

A barplot with `geom_bar()` with categorical variables

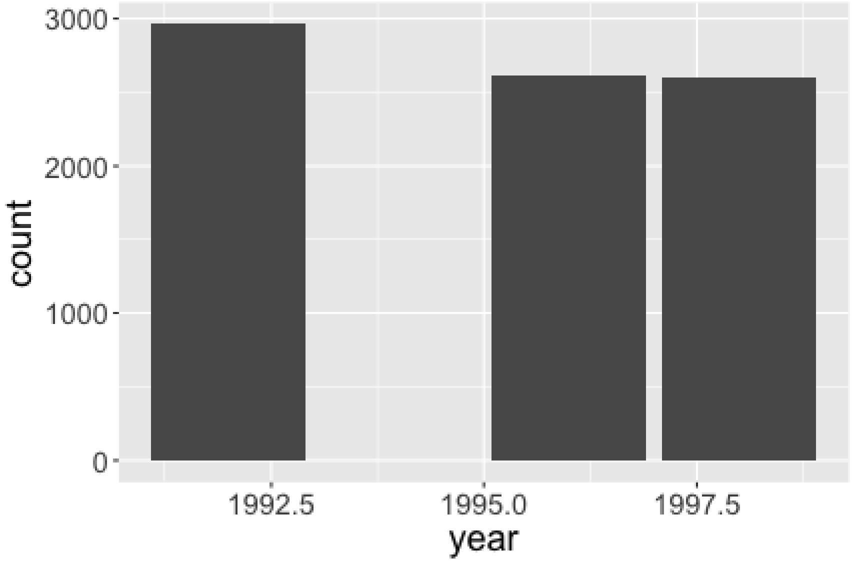
```
ggplot(data = CPSch3,
       mapping = aes(x = sex)) +
  geom_bar()
```



▸ Back to catalogue

A barplot with geom_bar( ) with discrete variables

```
ggplot(data = filter(CPSch3, year!=1994),
          mapping = aes(x = year)) +
   geom_bar()
```



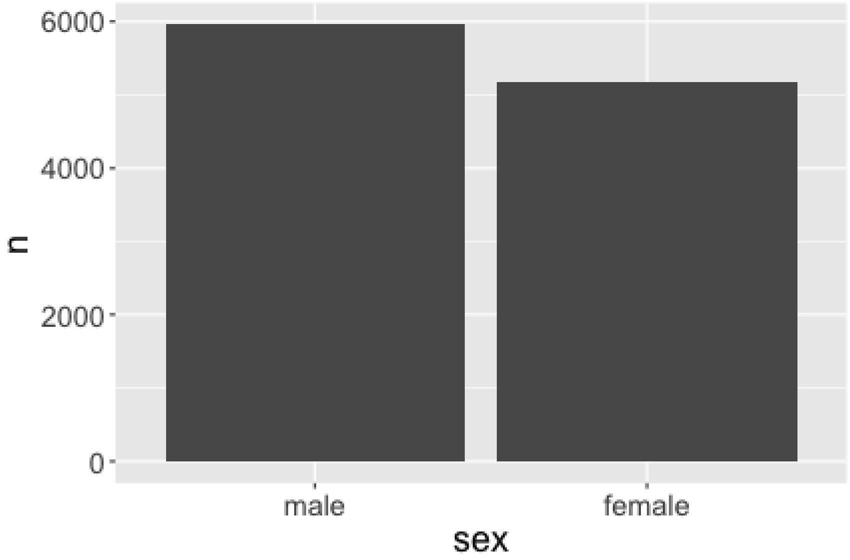
▸ Back to catalogue

A barplot with geom_col()

```
dfsum <- CPSch3 %>%
  group_by(sex) %>%
  tally()
dfsum
```

```
## # A tibble: 2 × 2
##   sex        n
##   <fct>  <int>
## 1 male    5956
## 2 female  5174
```

```
ggplot(data = dfsum,
       mapping = aes(x = sex,
                     y = n)) +
  geom_col()
```



Back to catalogue

Illusrative data 🚢 ship accidents

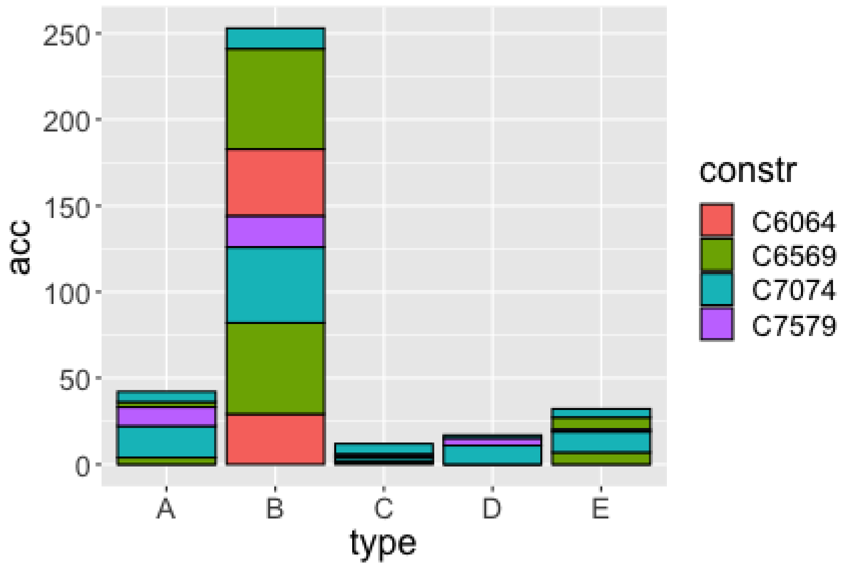
```r
data(Accident, package = "Ecdat")
str(Accident)
```

```
## 'data.frame':    40 obs. of  5 variables:
##  $ type    : Factor w/ 5 levels "A","B","C","D",..: 1 1 1
##  $ constr  : Factor w/ 4 levels "C6064","C6569",..: 1 1 2
##  $ operate: Factor w/ 2 levels "O6074","O7579": 1 2 1 2
##  $ months  : int  127 63 1095 1095 1512 3353 NA 2244 44882
##  $ acc     : int  0 0 3 4 6 18 NA 11 39 29 ...
```

▸ Back to catalogue

A stacked barplot with geom_col()

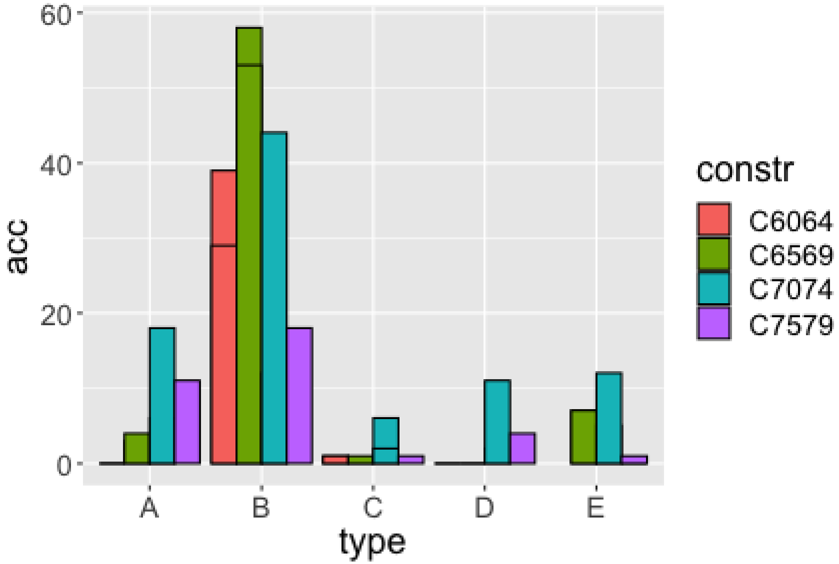
```
ggplot(Accident, aes(type, acc, fill = constr, group = opera
  geom_col(color = "black", position = "stack")
```



Back to catalogue

A grouped barplot with geom_col()

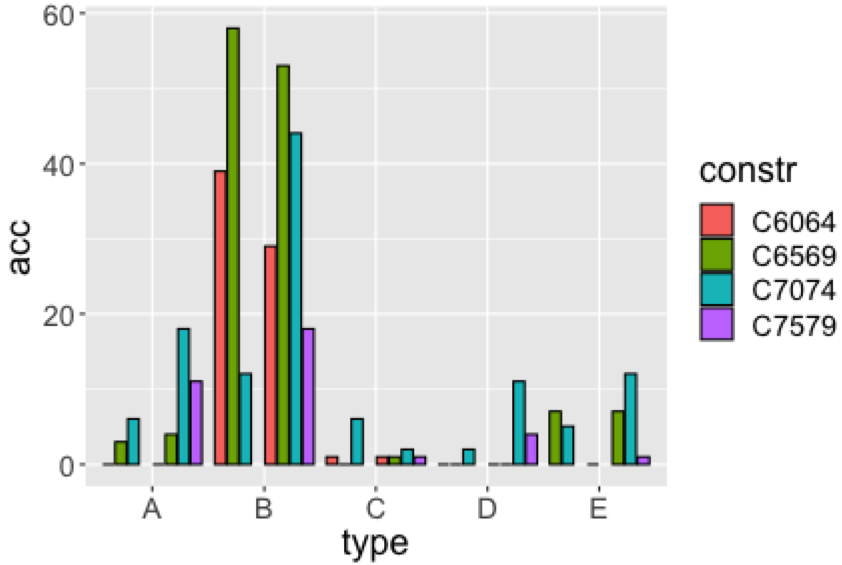
```
ggplot(Accident, aes(type, acc, fill = constr)) +
  geom_col(color = "black", position = "dodge")
```



▸ Back to catalogue

Another grouped barplot with geom_col()

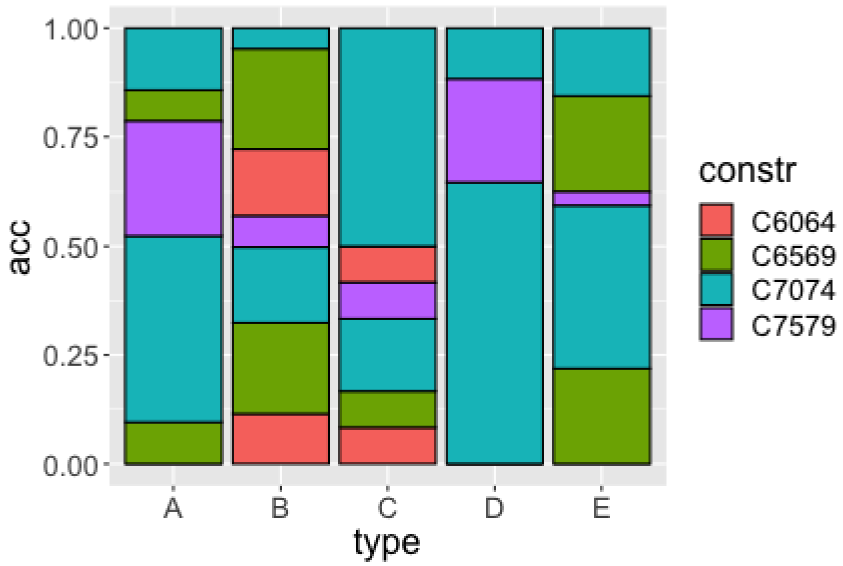
```
ggplot(Accident, aes(type, acc, fill = constr, group = opera
  geom_col(color = "black", position = "dodge2")
```



Back to catalogue

Stacked percentage barplot with geom_col()

```
ggplot(Accident, aes(type, acc, fill = constr, group = opera
  geom_col(color = "black", position = "fill")
```



▸ Back to catalogue

Illustrative data 💎 diamonds

- Let's have a look at the `diamonds` data which contains information about different attributes of diamonds

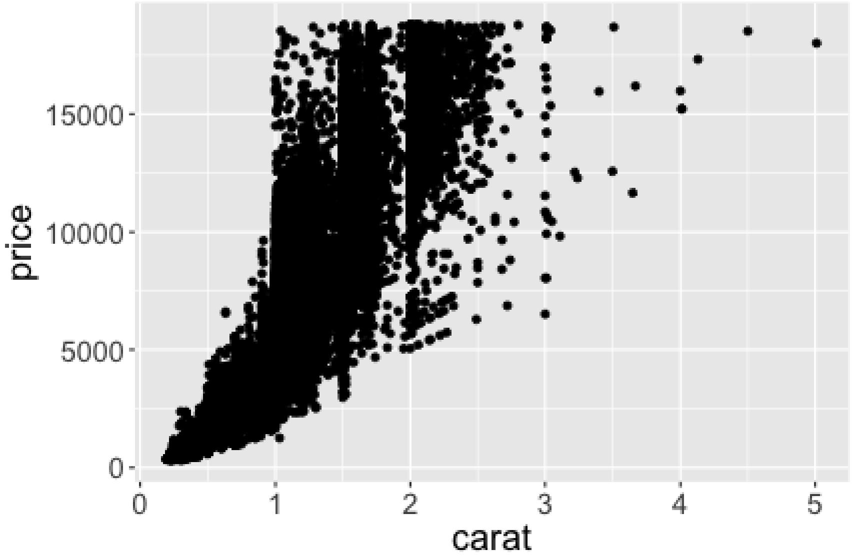
```
data(diamonds, package = "ggplot2")
diamonds
```

```
## # A tibble: 53,940 × 10
##    carat cut       color clarity depth table price     x
##    <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl>
##  1  0.23 Ideal     E     SI2      61.5    55   326  3.95
##  2  0.21 Premium   E     SI1      59.8    61   326  3.89
##  3  0.23 Good      E     VS1      56.9    65   327  4.05
##  4  0.29 Premium   I     VS2      62.4    58
```

▸ Back to catalogue

A scatterplot with geom_point()

```
ggplot(data = diamonds,
       mapping = aes(x = carat, y = price)) +
  geom_point()
```

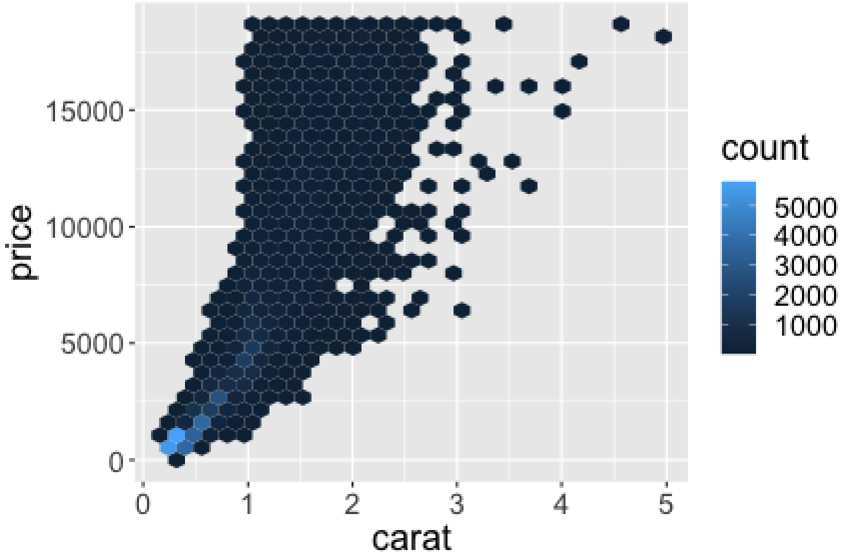


- Scatter plot of price vs carat of diamonds
- Each point correponds to a dimaond

Back to catalogue

A hexagonal 2D heatmap with geom_hex()

```
ggplot(data = diamonds,
       mapping = aes(x = carat, y = price)) +
  geom_hex()
```

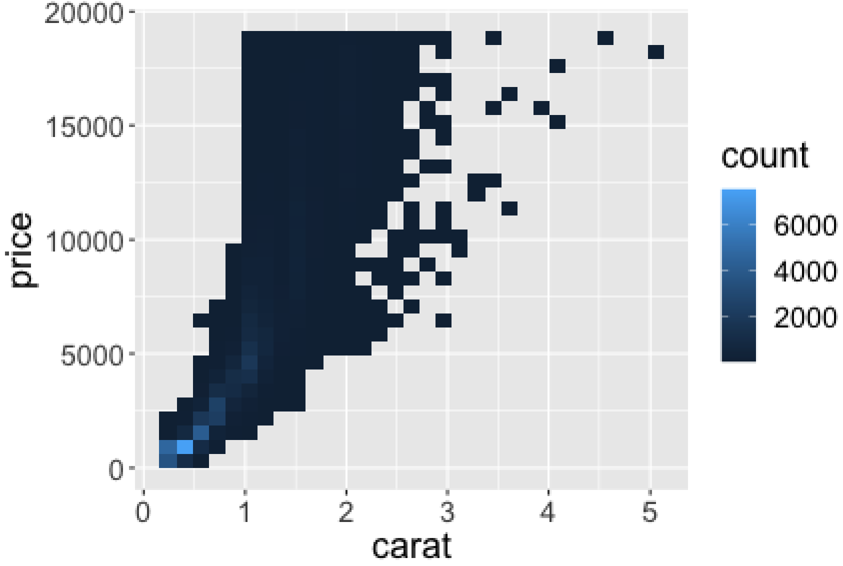


- A hexagon shows the count of observations within the region.

▸ Back to catalogue

A rectangular 2D heatmap with geom_bin_2d()

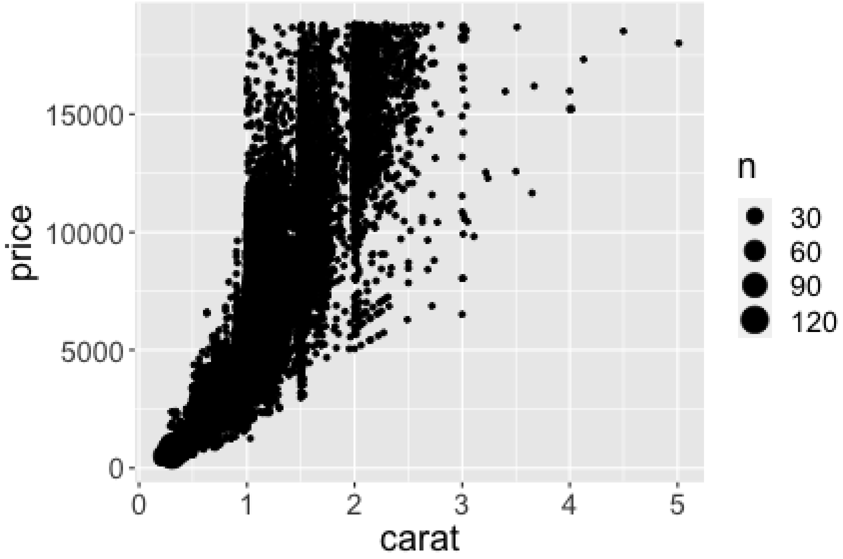
```
ggplot(data = diamonds,
       mapping = aes(x = carat, y = price)) +
  geom_bin_2d()
```



▸ Back to catalogue

Count of overlapping points with `geom_count()`

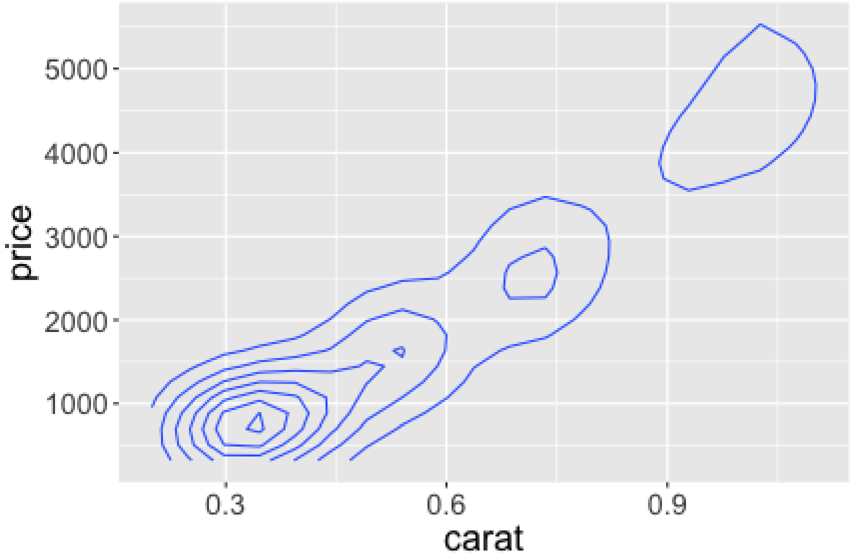
```
ggplot(data = diamonds,
       mapping = aes(x = carat, y = price)) +
  geom_count()
```



▸ Back to catalogue

Contour of a 2D density with `geom_density_2d()`

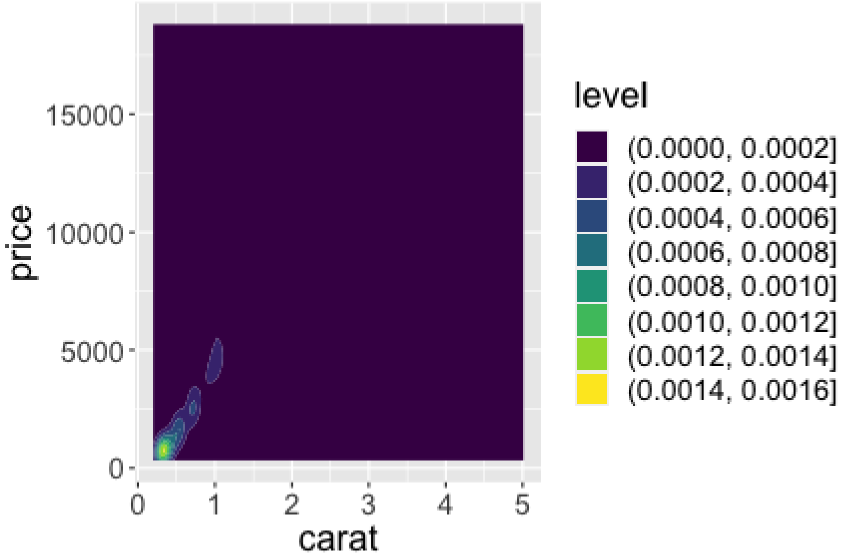
```
ggplot(data = diamonds,
       mapping = aes(x = carat, y = price)) +
  geom_density_2d()
```



▸ Back to catalogue

Contour of a 2D density with `geom_density_2d_filled()`

```
ggplot(data = diamonds,
       mapping = aes(x = carat, y = price)) +
  geom_density_2d_filled()
```



▸ Back to catalogue

Illustrative data ⛰️ volcano

- The data contains topographic information on Auckland's Maunga Whau Volcano

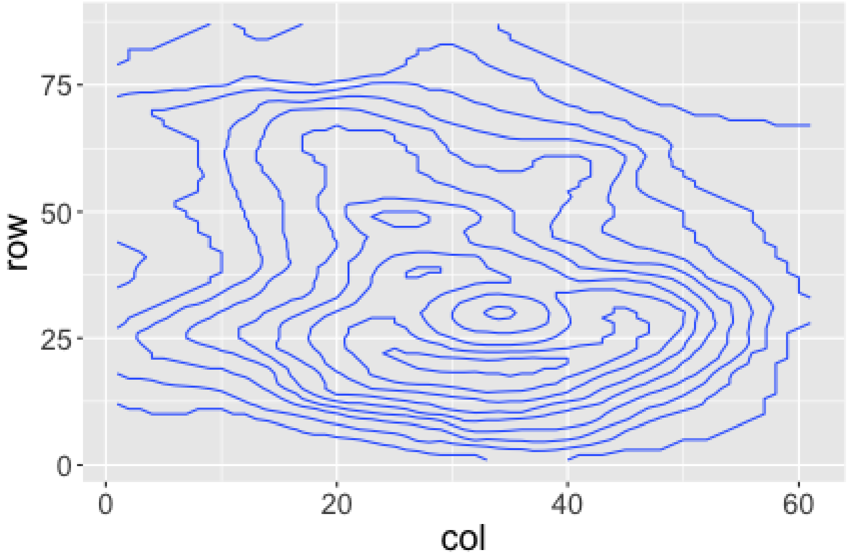
```
volcanod <- data.frame(row = as.vector(row(volcano)),
                       col = as.vector(col(volcano)),
                       value = as.vector(volcano))

volcanod
```

```
##       row col value
## 1       1   1   100
## 2       2   1   101
## 3       3   1   102
```

▸ Back to catalogue

A 2D contour plot with geom_contour()

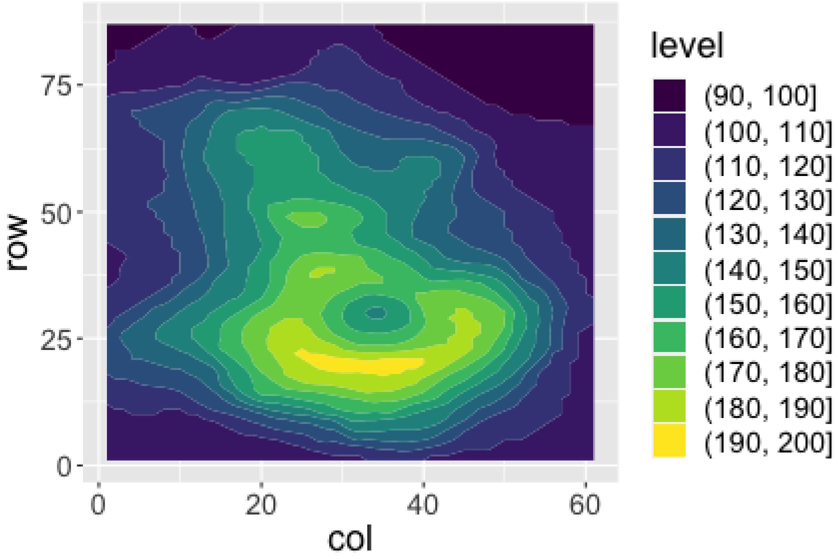
```
ggplot(data = volcanod,
       mapping = aes(x = col, y = row, z = value)) +
  geom_contour()
```



▸ Back to catalogue

A 2D contour plot with `geom_contour_filled()`

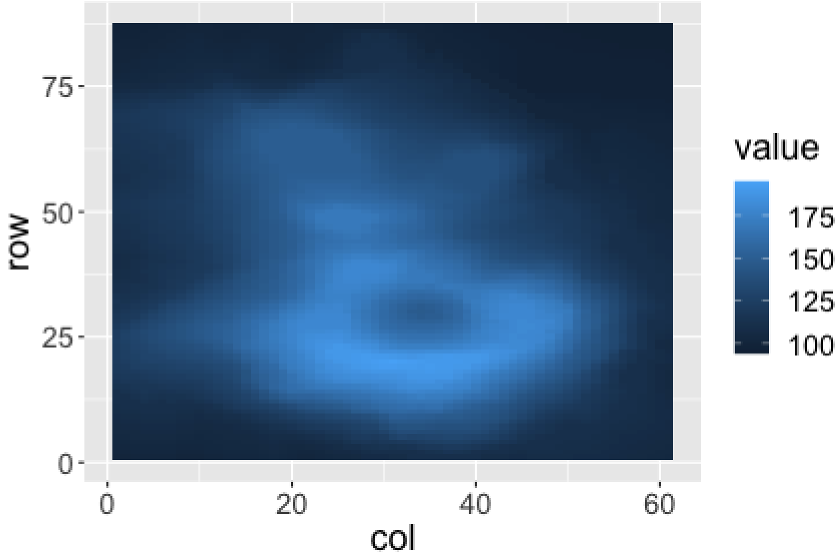
```
ggplot(data = volcanod,
       mapping = aes(x = col, y = row, z = value)) +
  geom_contour_filled()
```



▸ Back to catalogue

Tile plots with geom_tile()

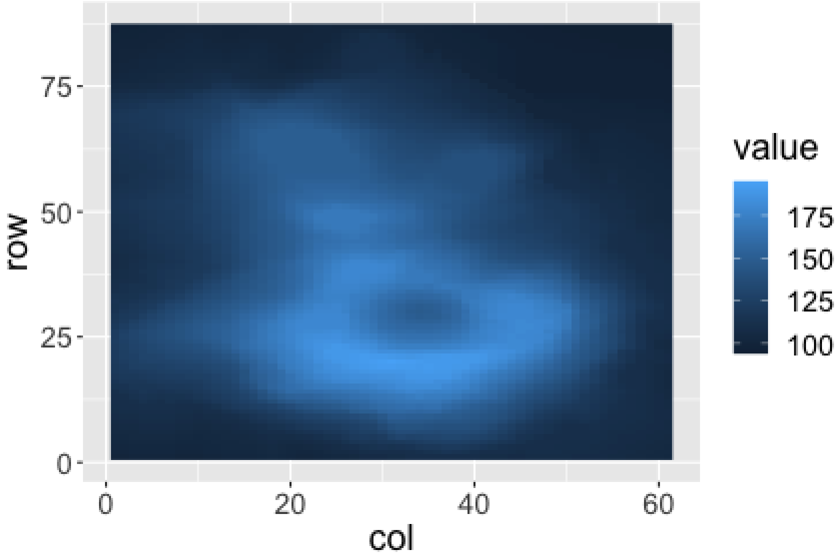
```
ggplot(volcanod, aes(col, row, fill = value)) +
 geom_tile()
```



Back to catalogue

Raster plots with `geom_raster()`

```
ggplot(volcanod, aes(col, row, fill = value)) +
 geom_raster()
```

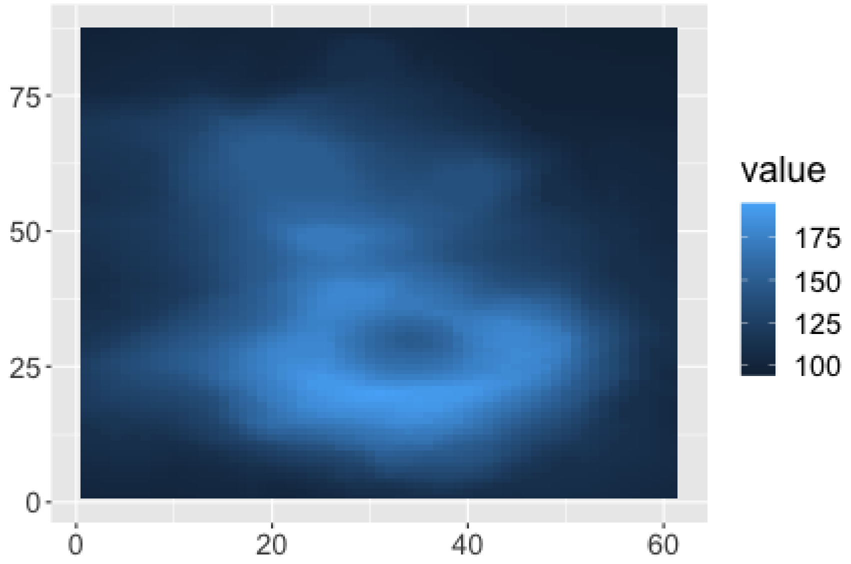


- A high performance special case of `geom_tile()` for when the tiles are the same size

▸ Back to catalogue

Rectangular plots with geom_rect()

```
ggplot(volcanod, aes(xmin = col - 0.5, xmax = col + 0.5,
                     ymin = row - 0.5, ymax = row + 0.5, fil
 geom_rect()
```



- A reparameterised version of `geom_tile()`

▸ Back to catalogue

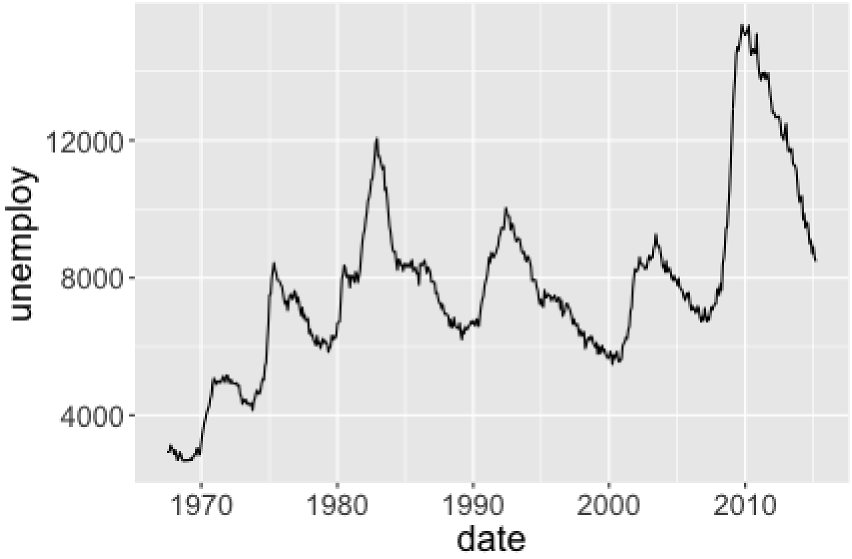
Illustrative data 📈 economics

```
data(economics, package = "ggplot2")
economics

## # A tibble: 574 × 6
##    date          pce    pop psavert uempmed unemploy
##    <date>      <dbl>  <dbl>   <dbl>   <dbl>    <dbl>
##  1 1967-07-01   507. 198712    12.6     4.5     2944
##  2 1967-08-01   510. 198911    12.6     4.7     2945
##  3 1967-09-01   516. 199113    11.9     4.6     2958
##  4 1967-10-01   512. 199311    12.9     4.9     3143
##  5 1967-11-01   517. 199498    12.8     4.7     3066
##  6 1967-12-01   525. 199657    11.8     4.8
```

▸ Back to catalogue

Line plot with geom_line()

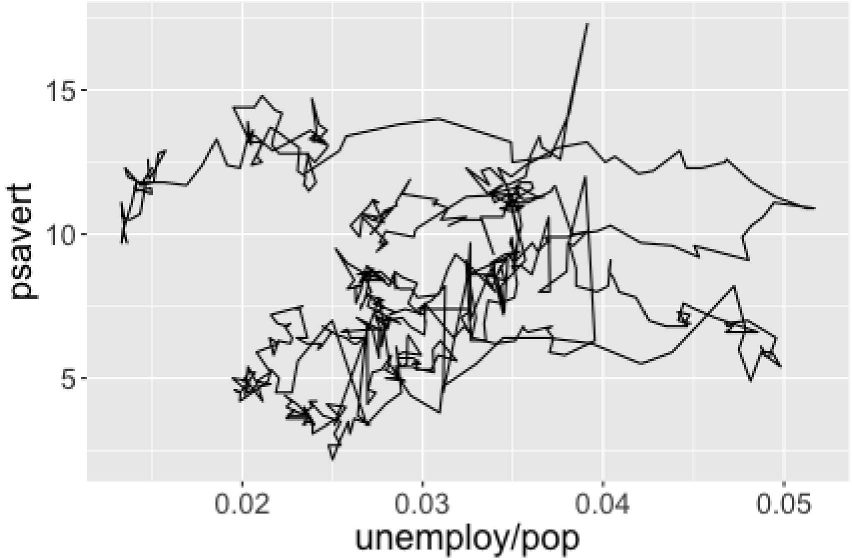
```
ggplot(economics, aes(date, unemploy)) +
  geom_line()
```



▸ Back to catalogue

geom_path()

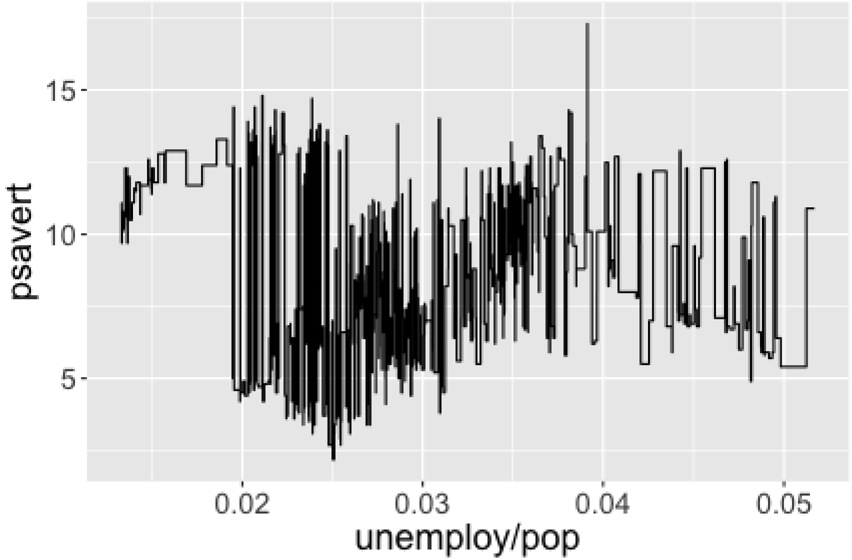
```
ggplot(economics, aes(unemploy/pop, psavert)) +
  geom_path()
```



▸ Back to catalogue

geom_step()

```
ggplot(economics, aes(unemploy/pop, psavert)) +
  geom_step()
```



▸ Back to catalogue

Drawing maps

- Drawing maps require the map data

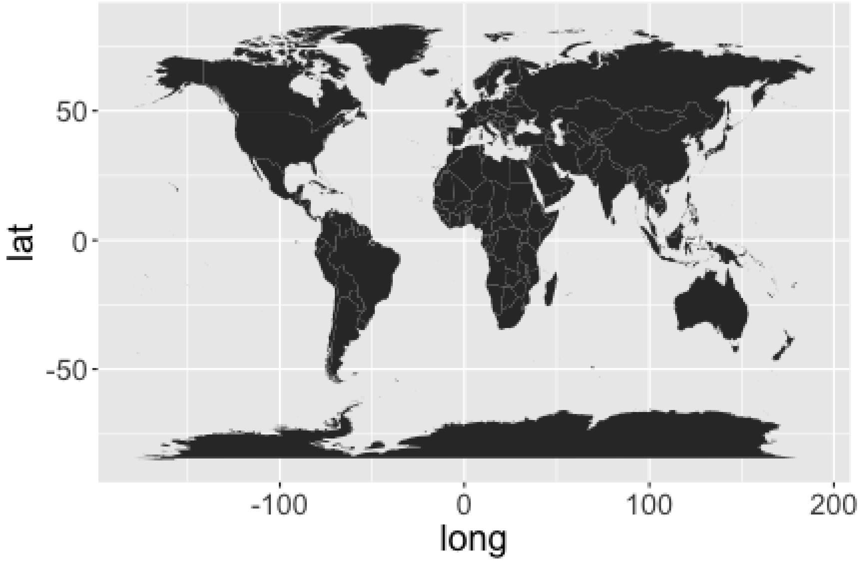
```
world <- map_data("world")
world
```

```
##          long      lat group order  region subregion
## 1   -69.89912 12.45200     1     1   Aruba      <NA>
## 2   -69.89571 12.42300     1     2   Aruba      <NA>
## 3   -69.94219 12.43853     1     3   Aruba      <NA>
## 4   -70.00415 12.50049     1     4   Aruba      <NA>
## 5   -70.06612 12.54697     1     5   Aruba      <NA>
## 6   -70.05088 12.59707     1     6   Aruba      <NA>
## 7   -70.03511 12.61411     1     7   Aruba
```

▸ Back to catalogue

Drawing world map with geom_polygon()

```
world <- map_data("world")
ggplot(world, aes(long, lat, group = group)) +
  geom_polygon()
```



▸ Back to catalogue

Other layers

- There are more layers in `ggplot`!
- There are layers also beginning with `stat_`.
- There are also many extension packages to draw more plots: https://exts.ggplot2.tidyverse.org/gallery/
- It is normal to google and copy-and-paste code that you find on the web.
- `ggplot2` is also continually updated so some new features may be added in the future.

▸ Back to catalogue

Session Information

```r
devtools::session_info()
```

```
## ─ Session info  3️⃣  🧛🏻  👨🏾  ──────────────────────────────────────────
##  hash: keycap: 3, vampire: light skin tone, man: medium-dark skin tone
##
##  setting  value
##  version  R version 4.1.2 (2021-11-01)
##  os       macOS Big Sur 11.6.2
##  system   x86_64, darwin17.0
##  ui       RStudio
##  language (EN)
##  collate  en_AU.UTF-8
##  ctype    en_AU.UTF-8
##  tz       Australia/Melbourne
##  date     2022-02-20
```

▸ Back to catalogue