Um breve tutorial sobre dplyr com ggplot2

true

10/03/2023

Sumário

Capítulo 1

Introdução

Palestra do *XIV Programa de Verão DES-ICET/UFLA - 2023*

1.1 Alguns atalhos no Rstudio

Para considerar

> Operador Pipe (%>%): Ctrl + Shift + M (Windows) ou Cmd + Shift + M (Mac).
>
> Criar novos chunks: Ctrl + Alt + I (Windows) ou Cmd + Option + I (Mac).

1.2 Carrega pacotes a serem usados

```
#install.packages("tidyverse")
#install.packages("dplyr")
#install.packages("tidyr")
#install.packages("ggplot2")

# Manipulação de dados
library(dplyr)

# Visualização de gráficos
library(ggplot2)
```

```
library(gridExtra)
library(patchwork)
library(plotly)
library(esquisse)

# Para dados gráfico de perfis
library(nlme)
```

1.3 Descrição dos dados `mpg`

Dados de economia de combustível de 1999 a 2008 para *38 modelos populares de carros*. Este conjunto de dados contém um subconjunto dos dados de economia de combustível que a EPA disponibiliza em *https://fueleconomy.gov/*. Ele contém apenas modelos que tiveram um novo lançamento a cada ano entre 1999 e 2008 - isso foi usado como um substituto para a popularidade do carro. Um *data frame* com 234 linhas e 11 variáveis:

- *manufacturer* nome do fabricante
- *model* nome do modelo
- *displ* cilindrada do motor, em litros
- *year* ano de fabricação
- *cyl* número de cilindros
- *trans* tipo de transmissão
- *drv* o tipo de trem de força, onde **f = tração dianteira**, **r = tração traseira** e **4 = 4wd**
- *cty* milhas urbanas por galão
- *hwy* milhas rodoviárias por galão
- *fl* tipo de combustível
- *class* "tipo" de carro

```
#help("mpg")
dados <- mpg
glimpse(dados)
```

```
## Rows: 234
## Columns: 11
## $ manufacturer <chr> "audi", "audi", "audi", "audi", "audi", "audi", "audi", "~
## $ model        <chr> "a4", "a4", "a4", "a4", "a4", "a4", "a4", "a4 quattro", "~
## $ displ        <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0, 2.~
## $ year         <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999, 200~
## $ cyl          <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 8, 8, ~
## $ trans        <chr> "auto(l5)", "manual(m5)", "manual(m6)", "auto(av)", "auto~
## $ drv          <chr> "f", "f", "f", "f", "f", "f", "f", "4", "4", "4", "4", "4~
## $ cty          <int> 18, 21, 20, 21, 16, 18, 18, 18, 16, 20, 19, 15, 17, 17, 1~
## $ hwy          <int> 29, 29, 31, 30, 26, 26, 27, 26, 25, 28, 27, 25, 25, 25, 2~
## $ fl           <chr> "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p~
## $ class        <chr> "compact", "compact", "compact", "compact", "compact", "c~
```

```
dados <- mutate(.data = dados,
                across(where(is.character),
                as.factor))
#View(df)
glimpse(dados)
```

```
## Rows: 234
## Columns: 11
## $ manufacturer <fct> audi, audi, audi, audi, audi, audi, audi, audi, audi, aud~
## $ model        <fct> a4, a4, a4, a4, a4, a4, a4, a4 quattro, a4 quattro, a4 qu~
## $ displ        <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0, 2.~
## $ year         <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999, 200~
## $ cyl          <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 8, 8, ~
## $ trans        <fct> auto(l5), manual(m5), manual(m6), auto(av), auto(l5), man~
## $ drv          <fct> f, f, f, f, f, f, f, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, r, ~
## $ cty          <int> 18, 21, 20, 21, 16, 18, 18, 18, 16, 20, 19, 15, 17, 17, 1~
## $ hwy          <int> 29, 29, 31, 30, 26, 26, 27, 26, 25, 28, 27, 25, 25, 25, 2~
## $ fl           <fct> p, p, p, p, p, p, p, p, p, p, p, p, p, p, p, p, p, p, r, ~
## $ class        <fct> compact, compact, compact, compact, compact, compact, com~
```

1.4 Propaganda 1 (Gustavo Jun Yakushiji)

Assistir os 3 vídeos do **Gustavo Jun Yakushiji** e **Cristian Villegas** no *youtube*

Dia 1 Introdução à Ciência de Dados (Introdução ao R e importação de dados)

Dia 2 Introdução à Ciência de Dados (Introdução ao ggplot2)

Dia 3 Introdução à Ciência de Dados (Introdução ao Tidyverse)

1.5 Propaganda 2 (Gustavo Jun Yakushiji)

Introdução à Ciência de Dados em R, Gustavo Jun Yakushiji; Cristian Villegas

1.6 Propaganda 3

Como pegar o código fonte do meu github? Link do meu github

Como pegar o arquivo PDF da apresentação? link: Um breve tutorial sobre dplyr com ggplot2 ou Arquivo PDF

Capítulo 2

dplyr (60 minutos)

2.1 Lista de funções do pacote dplyr

```
ls("package:dplyr")
```

```
##   [1] "%>%"                "across"             "add_count"
##   [4] "add_count_"         "add_row"            "add_rownames"
##   [7] "add_tally"          "add_tally_"         "all_equal"
##  [10] "all_of"             "all_vars"           "anti_join"
##  [13] "any_of"             "any_vars"           "arrange"
##  [16] "arrange_"           "arrange_all"        "arrange_at"
##  [19] "arrange_if"         "as.tbl"             "as_data_frame"
##  [22] "as_label"           "as_tibble"          "auto_copy"
##  [25] "band_instruments"   "band_instruments2"  "band_members"
##  [28] "bench_tbls"         "between"            "bind_cols"
##  [31] "bind_rows"          "c_across"           "case_when"
##  [34] "changes"            "check_dbplyr"       "coalesce"
##  [37] "collapse"           "collect"            "combine"
##  [40] "common_by"          "compare_tbls"       "compare_tbls2"
##  [43] "compute"            "contains"           "copy_to"
##  [46] "count"              "count_"             "cumall"
##  [49] "cumany"             "cume_dist"          "cummean"
##  [52] "cur_column"         "cur_data"           "cur_data_all"
##  [55] "cur_group"          "cur_group_id"       "cur_group_rows"
##  [58] "current_vars"       "data_frame"         "data_frame_"
##  [61] "db_analyze"         "db_begin"           "db_commit"
##  [64] "db_create_index"    "db_create_indexes"  "db_create_table"
##  [67] "db_data_type"       "db_desc"            "db_drop_table"
```

```
##  [70] "db_explain"            "db_has_table"          "db_insert_into"
##  [73] "db_list_tables"        "db_query_fields"       "db_query_rows"
##  [76] "db_rollback"           "db_save_query"         "db_write_table"
##  [79] "dense_rank"            "desc"                  "dim_desc"
##  [82] "distinct"              "distinct_"             "distinct_all"
##  [85] "distinct_at"           "distinct_if"           "distinct_prepare"
##  [88] "do"                    "do_"                   "dplyr_col_modify"
##  [91] "dplyr_reconstruct"     "dplyr_row_slice"       "ends_with"
##  [94] "enexpr"                "enexprs"               "enquo"
##  [97] "enquos"                "ensym"                 "ensyms"
## [100] "eval_tbls"             "eval_tbls2"            "everything"
## [103] "explain"               "expr"                  "failwith"
## [106] "filter"                "filter_"               "filter_all"
## [109] "filter_at"             "filter_if"             "first"
## [112] "frame_data"            "full_join"             "funs"
## [115] "funs_"                 "glimpse"               "group_by"
## [118] "group_by_"             "group_by_all"          "group_by_at"
## [121] "group_by_drop_default" "group_by_if"           "group_by_prepare"
## [124] "group_cols"            "group_data"            "group_indices"
## [127] "group_indices_"        "group_keys"            "group_map"
## [130] "group_modify"          "group_nest"            "group_rows"
## [133] "group_size"            "group_split"           "group_trim"
## [136] "group_vars"            "group_walk"            "grouped_df"
## [139] "groups"                "id"                    "ident"
## [142] "if_all"                "if_any"                "if_else"
## [145] "inner_join"            "intersect"             "is.grouped_df"
## [148] "is.src"                "is.tbl"                "is_grouped_df"
## [151] "lag"                   "last"                  "last_col"
## [154] "lead"                  "left_join"             "location"
## [157] "lst"                   "lst_"                  "make_tbl"
## [160] "matches"               "min_rank"              "mutate"
## [163] "mutate_"               "mutate_all"            "mutate_at"
## [166] "mutate_each"           "mutate_each_"          "mutate_if"
## [169] "n"                     "n_distinct"            "n_groups"
## [172] "na_if"                 "near"                  "nest_by"
## [175] "nest_join"             "new_grouped_df"        "new_rowwise_df"
## [178] "nth"                   "ntile"                 "num_range"
## [181] "one_of"                "order_by"              "percent_rank"
## [184] "progress_estimated"    "pull"                  "quo"
## [187] "quo_name"              "quos"                  "recode"
## [190] "recode_factor"         "relocate"              "rename"
## [193] "rename_"               "rename_all"            "rename_at"
## [196] "rename_if"             "rename_vars"           "rename_vars_"
## [199] "rename_with"           "right_join"            "row_number"
## [202] "rows_append"           "rows_delete"           "rows_insert"
## [205] "rows_patch"            "rows_update"           "rows_upsert"
```

```
## [208] "rowwise"              "same_src"              "sample_frac"
## [211] "sample_n"             "select"                "select_"
## [214] "select_all"           "select_at"             "select_if"
## [217] "select_var"           "select_vars"           "select_vars_"
## [220] "semi_join"            "setdiff"               "setequal"
## [223] "show_query"           "slice"                 "slice_"
## [226] "slice_head"           "slice_max"             "slice_min"
## [229] "slice_sample"         "slice_tail"            "sql"
## [232] "sql_escape_ident"     "sql_escape_string"     "sql_join"
## [235] "sql_select"           "sql_semi_join"         "sql_set_op"
## [238] "sql_subquery"         "sql_translate_env"     "src"
## [241] "src_df"               "src_local"             "src_mysql"
## [244] "src_postgres"         "src_sqlite"            "src_tbls"
## [247] "starts_with"          "starwars"              "storms"
## [250] "summarise"            "summarise_"            "summarise_all"
## [253] "summarise_at"         "summarise_each"        "summarise_each_"
## [256] "summarise_if"         "summarize"             "summarize_"
## [259] "summarize_all"        "summarize_at"          "summarize_each"
## [262] "summarize_each_"      "summarize_if"          "sym"
## [265] "syms"                 "tally"                 "tally_"
## [268] "tbl"                  "tbl_df"                "tbl_nongroup_vars"
## [271] "tbl_ptype"            "tbl_sum"               "tbl_vars"
## [274] "tibble"               "top_frac"              "top_n"
## [277] "transmute"            "transmute_"            "transmute_all"
## [280] "transmute_at"         "transmute_if"          "tribble"
## [283] "type_sum"             "ungroup"               "union"
## [286] "union_all"            "validate_grouped_df"   "validate_rowwise_df"
## [289] "vars"                 "with_groups"           "with_order"
## [292] "wrap_dbplyr_obj"
```

2.2 Operador Pipe

```
sqrt(log(44))
```

```
## [1] 1.945299
```

```
44 %>% log %>% sqrt
```

```
## [1] 1.945299
```

2.3 select() para colunas

```
select(dados, manufacturer, model, year)
```

```
## # A tibble: 234 x 3
##    manufacturer model       year
##    <fct>        <fct>      <int>
##  1 audi         a4          1999
##  2 audi         a4          1999
##  3 audi         a4          2008
##  4 audi         a4          2008
##  5 audi         a4          1999
##  6 audi         a4          1999
##  7 audi         a4          2008
##  8 audi         a4 quattro  1999
##  9 audi         a4 quattro  1999
## 10 audi         a4 quattro  2008
## # ... with 224 more rows
```

```
select(dados, starts_with("m"))
```

```
## # A tibble: 234 x 2
##    manufacturer model
##    <fct>        <fct>
##  1 audi         a4
##  2 audi         a4
##  3 audi         a4
##  4 audi         a4
##  5 audi         a4
##  6 audi         a4
##  7 audi         a4
##  8 audi         a4 quattro
##  9 audi         a4 quattro
## 10 audi         a4 quattro
## # ... with 224 more rows
```

```
select(dados, contains("r"))
```

```
## # A tibble: 234 x 4
##    manufacturer  year trans      drv
##    <fct>        <int> <fct>      <fct>
##  1 audi          1999 auto(l5)   f
```

```
##  2 audi          1999 manual(m5) f
##  3 audi          2008 manual(m6) f
##  4 audi          2008 auto(av)   f
##  5 audi          1999 auto(l5)   f
##  6 audi          1999 manual(m5) f
##  7 audi          2008 auto(av)   f
##  8 audi          1999 manual(m5) 4
##  9 audi          1999 auto(l5)   4
## 10 audi          2008 manual(m6) 4
## # ... with 224 more rows
```

```
select(dados, ends_with("y"))
```

```
## # A tibble: 234 x 2
##      cty   hwy
##    <int> <int>
##  1    18    29
##  2    21    29
##  3    20    31
##  4    21    30
##  5    16    26
##  6    18    26
##  7    18    27
##  8    18    26
##  9    16    25
## 10    20    28
## # ... with 224 more rows
```

```
select(dados, matches("[abc]"))
```

```
## # A tibble: 234 x 6
##    manufacturer  year   cyl trans        cty class
##    <fct>        <int> <int> <fct>      <int> <fct>
##  1 audi          1999     4 auto(l5)      18 compact
##  2 audi          1999     4 manual(m5)    21 compact
##  3 audi          2008     4 manual(m6)    20 compact
##  4 audi          2008     4 auto(av)      21 compact
##  5 audi          1999     6 auto(l5)      16 compact
##  6 audi          1999     6 manual(m5)    18 compact
##  7 audi          2008     6 auto(av)      18 compact
##  8 audi          1999     4 manual(m5)    18 compact
##  9 audi          1999     4 auto(l5)      16 compact
## 10 audi          2008     4 manual(m6)    20 compact
## # ... with 224 more rows
```

```
select(dados, starts_with("m"), starts_with("c"))
```

```
## # A tibble: 234 x 5
##    manufacturer model        cyl   cty class
##    <fct>        <fct>      <int> <int> <fct>
##  1 audi         a4             4    18 compact
##  2 audi         a4             4    21 compact
##  3 audi         a4             4    20 compact
##  4 audi         a4             4    21 compact
##  5 audi         a4             6    16 compact
##  6 audi         a4             6    18 compact
##  7 audi         a4             6    18 compact
##  8 audi         a4 quattro     4    18 compact
##  9 audi         a4 quattro     4    16 compact
## 10 audi         a4 quattro     4    20 compact
## # ... with 224 more rows
```

```
select(dados, ends_with("l"), ends_with("s"))
```

```
## # A tibble: 234 x 6
##    model      displ   cyl fl    trans      class
##    <fct>      <dbl> <int> <fct> <fct>      <fct>
##  1 a4           1.8     4 p     auto(l5)   compact
##  2 a4           1.8     4 p     manual(m5) compact
##  3 a4           2       4 p     manual(m6) compact
##  4 a4           2       4 p     auto(av)   compact
##  5 a4           2.8     6 p     auto(l5)   compact
##  6 a4           2.8     6 p     manual(m5) compact
##  7 a4           3.1     6 p     auto(av)   compact
##  8 a4 quattro   1.8     4 p     manual(m5) compact
##  9 a4 quattro   1.8     4 p     auto(l5)   compact
## 10 a4 quattro   2       4 p     manual(m6) compact
## # ... with 224 more rows
```

```
select(dados, 1:3)
```

```
## # A tibble: 234 x 3
##    manufacturer model      displ
##    <fct>        <fct>      <dbl>
##  1 audi         a4           1.8
##  2 audi         a4           1.8
##  3 audi         a4           2
##  4 audi         a4           2
```

```
##  5 audi         a4           2.8
##  6 audi         a4           2.8
##  7 audi         a4           3.1
##  8 audi         a4 quattro   1.8
##  9 audi         a4 quattro   1.8
## 10 audi         a4 quattro   2
## # ... with 224 more rows
```

```
select(dados, c(2,5,7))
```

```
## # A tibble: 234 x 3
##    model        cyl drv
##    <fct>      <int> <fct>
##  1 a4             4 f
##  2 a4             4 f
##  3 a4             4 f
##  4 a4             4 f
##  5 a4             6 f
##  6 a4             6 f
##  7 a4             6 f
##  8 a4 quattro     4 4
##  9 a4 quattro     4 4
## 10 a4 quattro     4 4
## # ... with 224 more rows
```

```
select(dados, manufacturer:cyl)
```

```
## # A tibble: 234 x 5
##    manufacturer model      displ  year   cyl
##    <fct>        <fct>      <dbl> <int> <int>
##  1 audi         a4           1.8  1999     4
##  2 audi         a4           1.8  1999     4
##  3 audi         a4           2    2008     4
##  4 audi         a4           2    2008     4
##  5 audi         a4           2.8  1999     6
##  6 audi         a4           2.8  1999     6
##  7 audi         a4           3.1  2008     6
##  8 audi         a4 quattro   1.8  1999     4
##  9 audi         a4 quattro   1.8  1999     4
## 10 audi         a4 quattro   2    2008     4
## # ... with 224 more rows
```

```
select(dados,-(manufacturer:cyl))
```

```
## # A tibble: 234 x 6
##    trans      drv     cty   hwy fl    class
##    <fct>      <fct> <int> <int> <fct> <fct>
##  1 auto(l5)   f        18    29 p     compact
##  2 manual(m5) f        21    29 p     compact
##  3 manual(m6) f        20    31 p     compact
##  4 auto(av)   f        21    30 p     compact
##  5 auto(l5)   f        16    26 p     compact
##  6 manual(m5) f        18    26 p     compact
##  7 auto(av)   f        18    27 p     compact
##  8 manual(m5) 4        18    26 p     compact
##  9 auto(l5)   4        16    25 p     compact
## 10 manual(m6) 4        20    28 p     compact
## # ... with 224 more rows
```

2.4 rename()

```
dados1 <- rename(dados,
                 mnfc = manufacturer,
                 mod = model)
dados1
```

```
## # A tibble: 234 x 11
##    mnfc  mod        displ  year   cyl trans      drv     cty   hwy fl    class
##    <fct> <fct>      <dbl> <int> <int> <fct>      <fct> <int> <int> <fct> <fct>
##  1 audi  a4           1.8  1999     4 auto(l5)   f        18    29 p     compact
##  2 audi  a4           1.8  1999     4 manual(m5) f        21    29 p     compact
##  3 audi  a4           2    2008     4 manual(m6) f        20    31 p     compact
##  4 audi  a4           2    2008     4 auto(av)   f        21    30 p     compact
##  5 audi  a4           2.8  1999     6 auto(l5)   f        16    26 p     compact
##  6 audi  a4           2.8  1999     6 manual(m5) f        18    26 p     compact
##  7 audi  a4           3.1  2008     6 auto(av)   f        18    27 p     compact
##  8 audi  a4 quattro   1.8  1999     4 manual(m5) 4        18    26 p     compact
##  9 audi  a4 quattro   1.8  1999     4 auto(l5)   4        16    25 p     compact
## 10 audi  a4 quattro   2    2008     4 manual(m6) 4        20    28 p     compact
## # ... with 224 more rows
```

```
select(dados,
       mnfc = manufacturer,
       mod = model)
```

```
## # A tibble: 234 x 2
##    mnfc  mod
##    <fct> <fct>
##  1 audi  a4
##  2 audi  a4
##  3 audi  a4
##  4 audi  a4
##  5 audi  a4
##  6 audi  a4
##  7 audi  a4
##  8 audi  a4 quattro
##  9 audi  a4 quattro
## 10 audi  a4 quattro
## # ... with 224 more rows
```

```
select(dados,
       mnfc = manufacturer,
       mod = model,
       everything())
```

```
## # A tibble: 234 x 11
##    mnfc  mod        displ  year   cyl trans      drv     cty   hwy fl    class
##    <fct> <fct>      <dbl> <int> <int> <fct>      <fct> <int> <int> <fct> <fct>
##  1 audi  a4           1.8  1999     4 auto(l5)   f        18    29 p     compact
##  2 audi  a4           1.8  1999     4 manual(m5) f        21    29 p     compact
##  3 audi  a4           2    2008     4 manual(m6) f        20    31 p     compact
##  4 audi  a4           2    2008     4 auto(av)   f        21    30 p     compact
##  5 audi  a4           2.8  1999     6 auto(l5)   f        16    26 p     compact
##  6 audi  a4           2.8  1999     6 manual(m5) f        18    26 p     compact
##  7 audi  a4           3.1  2008     6 auto(av)   f        18    27 p     compact
##  8 audi  a4 quattro   1.8  1999     4 manual(m5) 4        18    26 p     compact
##  9 audi  a4 quattro   1.8  1999     4 auto(l5)   4        16    25 p     compact
## 10 audi  a4 quattro   2    2008     4 manual(m6) 4        20    28 p     compact
## # ... with 224 more rows
```

2.5 mutate() para colunas

```
mutate(dados, sqrt_cty = sqrt(cty))
```

```
## # A tibble: 234 x 12
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
```

```
##  1 audi      a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
##  2 audi      a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  3 audi      a4      2    2008     4 manu~ f        20    31 p     comp~    4.47
##  4 audi      a4      2    2008     4 auto~ f        21    30 p     comp~    4.58
##  5 audi      a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
##  6 audi      a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
##  7 audi      a4      3.1  2008     6 auto~ f        18    27 p     comp~    4.24
##  8 audi      a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi      a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 audi      a4 q~   2    2008     4 manu~ 4        20    28 p     comp~    4.47
## # ... with 224 more rows, and abbreviated variable names 1: manufacturer,
## #   2: sqrt_cty
```

```
names(dados)
```

```
##  [1] "manufacturer" "model"        "displ"        "year"         "cyl"
##  [6] "trans"        "drv"          "cty"          "hwy"          "fl"
## [11] "class"
```

```
dados<- mutate(dados, sqrt_cty = sqrt(cty))
names(dados)
```

```
##  [1] "manufacturer" "model"        "displ"        "year"         "cyl"
##  [6] "trans"        "drv"          "cty"          "hwy"          "fl"
## [11] "class"        "sqrt_cty"
```

```
dados <- mutate(dados,
`soma de variáveis` = (cty + hwy) / 2)
names(dados)
```

```
##  [1] "manufacturer"      "model"             "displ"
##  [4] "year"              "cyl"               "trans"
##  [7] "drv"               "cty"               "hwy"
## [10] "fl"                "class"             "sqrt_cty"
## [13] "soma de variáveis"
```

```
dados <- mutate(dados,
            car = paste(manufacturer, model, sep = " "),
            `cyl / trans` = paste(cyl, " cylinders", " / ", trans, " transmission", se
dados
```

```
## # A tibble: 234 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
```

```
##    <fct>      <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 audi       a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
##  2 audi       a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  3 audi       a4      2    2008     4 manu~ f        20    31 p     comp~    4.47
##  4 audi       a4      2    2008     4 auto~ f        21    30 p     comp~    4.58
##  5 audi       a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
##  6 audi       a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
##  7 audi       a4      3.1  2008     6 auto~ f        18    27 p     comp~    4.24
##  8 audi       a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi       a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 audi       a4 q~   2    2008     4 manu~ 4        20    28 p     comp~    4.47
## # ... with 224 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

2.6 transmute()

```
transmute(dados,
          `avg miles per gallon` = (cty + hwy) / 2)
```

```
## # A tibble: 234 x 1
##    `avg miles per gallon`
##                     <dbl>
##  1                   23.5
##  2                   25
##  3                   25.5
##  4                   25.5
##  5                   21
##  6                   22
##  7                   22.5
##  8                   22
##  9                   20.5
## 10                   24
## # ... with 224 more rows
```

```
transmute(dados,
          car = paste(manufacturer, model, sep = " "),
          `cyl / trans` = paste(cyl, " cylinders", " / ", trans, " transmission", sep = ""))
```

```
## # A tibble: 234 x 2
##    car             `cyl / trans`
##    <chr>           <chr>
```

```
##  1 audi a4         4 cylinders / auto(l5) transmission
##  2 audi a4         4 cylinders / manual(m5) transmission
##  3 audi a4         4 cylinders / manual(m6) transmission
##  4 audi a4         4 cylinders / auto(av) transmission
##  5 audi a4         6 cylinders / auto(l5) transmission
##  6 audi a4         6 cylinders / manual(m5) transmission
##  7 audi a4         6 cylinders / auto(av) transmission
##  8 audi a4 quattro 4 cylinders / manual(m5) transmission
##  9 audi a4 quattro 4 cylinders / auto(l5) transmission
## 10 audi a4 quattro 4 cylinders / manual(m6) transmission
## # ... with 224 more rows
```

2.7 filter() para linhas

```
filter(dados, manufacturer == "audi")
```

```
## # A tibble: 18 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 audi      a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
##  2 audi      a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  3 audi      a4      2    2008     4 manu~ f        20    31 p     comp~    4.47
##  4 audi      a4      2    2008     4 auto~ f        21    30 p     comp~    4.58
##  5 audi      a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
##  6 audi      a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
##  7 audi      a4      3.1  2008     6 auto~ f        18    27 p     comp~    4.24
##  8 audi      a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi      a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 audi      a4 q~   2    2008     4 manu~ 4        20    28 p     comp~    4.47
## 11 audi      a4 q~   2    2008     4 auto~ 4        19    27 p     comp~    4.36
## 12 audi      a4 q~   2.8  1999     6 auto~ 4        15    25 p     comp~    3.87
## 13 audi      a4 q~   2.8  1999     6 manu~ 4        17    25 p     comp~    4.12
## 14 audi      a4 q~   3.1  2008     6 auto~ 4        17    25 p     comp~    4.12
## 15 audi      a4 q~   3.1  2008     6 manu~ 4        15    25 p     comp~    3.87
## 16 audi      a6 q~   2.8  1999     6 auto~ 4        15    24 p     mids~    3.87
## 17 audi      a6 q~   3.1  2008     6 auto~ 4        17    25 p     mids~    4.12
## 18 audi      a6 q~   4.2  2008     8 auto~ 4        16    23 p     mids~    4
## # ... with 3 more variables: `soma de variáveis` <dbl>, car <chr>,
## #   `cyl / trans` <chr>, and abbreviated variable names 1: manufacturer,
## #   2: sqrt_cty
```

```
filter(dados, manufacturer == "audi" & year == "1999")
```

```
## # A tibble: 9 x 15
##   manufact~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##   <fct>      <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
## 1 audi       a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
## 2 audi       a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
## 3 audi       a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
## 4 audi       a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
## 5 audi       a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
## 6 audi       a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 7 audi       a4 q~   2.8  1999     6 auto~ 4        15    25 p     comp~    3.87
## 8 audi       a4 q~   2.8  1999     6 manu~ 4        17    25 p     comp~    4.12
## 9 audi       a6 q~   2.8  1999     6 auto~ 4        15    24 p     mids~    3.87
## # ... with 3 more variables: `soma de variáveis` <dbl>, car <chr>,
## #   `cyl / trans` <chr>, and abbreviated variable names 1: manufacturer,
## #   2: sqrt_cty
```

```
filter(dados, manufacturer == "audi", year == 1999)
```

```
## # A tibble: 9 x 15
##   manufact~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##   <fct>      <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
## 1 audi       a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
## 2 audi       a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
## 3 audi       a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
## 4 audi       a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
## 5 audi       a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
## 6 audi       a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 7 audi       a4 q~   2.8  1999     6 auto~ 4        15    25 p     comp~    3.87
## 8 audi       a4 q~   2.8  1999     6 manu~ 4        17    25 p     comp~    4.12
## 9 audi       a6 q~   2.8  1999     6 auto~ 4        15    24 p     mids~    3.87
## # ... with 3 more variables: `soma de variáveis` <dbl>, car <chr>,
## #   `cyl / trans` <chr>, and abbreviated variable names 1: manufacturer,
## #   2: sqrt_cty
```

```
filter(dados, manufacturer == "audi" | manufacturer == "dodge") %>%
  print(n = 20)
```

```
## # A tibble: 55 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 audi      a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
```

```
##  2 audi      a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  3 audi      a4      2    2008     4 manu~ f        20    31 p     comp~    4.47
##  4 audi      a4      2    2008     4 auto~ f        21    30 p     comp~    4.58
##  5 audi      a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
##  6 audi      a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
##  7 audi      a4      3.1  2008     6 auto~ f        18    27 p     comp~    4.24
##  8 audi      a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi      a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 audi      a4 q~   2    2008     4 manu~ 4        20    28 p     comp~    4.47
## 11 audi      a4 q~   2    2008     4 auto~ 4        19    27 p     comp~    4.36
## 12 audi      a4 q~   2.8  1999     6 auto~ 4        15    25 p     comp~    3.87
## 13 audi      a4 q~   2.8  1999     6 manu~ 4        17    25 p     comp~    4.12
## 14 audi      a4 q~   3.1  2008     6 auto~ 4        17    25 p     comp~    4.12
## 15 audi      a4 q~   3.1  2008     6 manu~ 4        15    25 p     comp~    3.87
## 16 audi      a6 q~   2.8  1999     6 auto~ 4        15    24 p     mids~    3.87
## 17 audi      a6 q~   3.1  2008     6 auto~ 4        17    25 p     mids~    4.12
## 18 audi      a6 q~   4.2  2008     8 auto~ 4        16    23 p     mids~    4
## 19 dodge     cara~   2.4  1999     4 auto~ f        18    24 r     mini~    4.24
## 20 dodge     cara~   3    1999     6 auto~ f        17    24 r     mini~    4.12
## # ... with 35 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

```
filter(dados, manufacturer %in% c("audi", "dodge")) %>%
  print(n = 20)
```

```
## # A tibble: 55 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 audi      a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
##  2 audi      a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  3 audi      a4      2    2008     4 manu~ f        20    31 p     comp~    4.47
##  4 audi      a4      2    2008     4 auto~ f        21    30 p     comp~    4.58
##  5 audi      a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
##  6 audi      a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
##  7 audi      a4      3.1  2008     6 auto~ f        18    27 p     comp~    4.24
##  8 audi      a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi      a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 audi      a4 q~   2    2008     4 manu~ 4        20    28 p     comp~    4.47
## 11 audi      a4 q~   2    2008     4 auto~ 4        19    27 p     comp~    4.36
## 12 audi      a4 q~   2.8  1999     6 auto~ 4        15    25 p     comp~    3.87
## 13 audi      a4 q~   2.8  1999     6 manu~ 4        17    25 p     comp~    4.12
## 14 audi      a4 q~   3.1  2008     6 auto~ 4        17    25 p     comp~    4.12
## 15 audi      a4 q~   3.1  2008     6 manu~ 4        15    25 p     comp~    3.87
## 16 audi      a6 q~   2.8  1999     6 auto~ 4        15    24 p     mids~    3.87
```

```
## 17 audi      a6 q~   3.1  2008     6 auto~ 4        17    25 p     mids~    4.12
## 18 audi      a6 q~   4.2  2008     8 auto~ 4        16    23 p     mids~    4
## 19 dodge     cara~   2.4  1999     4 auto~ f        18    24 r     mini~    4.24
## 20 dodge     cara~   3    1999     6 auto~ f        17    24 r     mini~    4.12
## # ... with 35 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

```
filter(dados, hwy >= 30) %>%
  select(hwy) %>%
  print(n = 26)
```

```
## # A tibble: 26 x 1
##      hwy
##    <int>
##  1    31
##  2    30
##  3    30
##  4    33
##  5    32
##  6    32
##  7    32
##  8    34
##  9    36
## 10    36
## 11    30
## 12    31
## 13    31
## 14    32
## 15    31
## 16    31
## 17    31
## 18    31
## 19    30
## 20    33
## 21    35
## 22    37
## 23    35
## 24    44
## 25    44
## 26    41
```

```
filter(dados, year != 1999) %>%
  select(year) %>%
  print(n = 30)
```

```
## # A tibble: 117 x 1
##     year
##    <int>
##  1  2008
##  2  2008
##  3  2008
##  4  2008
##  5  2008
##  6  2008
##  7  2008
##  8  2008
##  9  2008
## 10  2008
## 11  2008
## 12  2008
## 13  2008
## 14  2008
## 15  2008
## 16  2008
## 17  2008
## 18  2008
## 19  2008
## 20  2008
## 21  2008
## 22  2008
## 23  2008
## 24  2008
## 25  2008
## 26  2008
## 27  2008
## 28  2008
## 29  2008
## 30  2008
## # ... with 87 more rows
```

```
filter(dados, between(cty,15, 22))
```

```
## # A tibble: 143 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 audi      a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
##  2 audi      a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  3 audi      a4      2    2008     4 manu~ f        20    31 p     comp~    4.47
##  4 audi      a4      2    2008     4 auto~ f        21    30 p     comp~    4.58
##  5 audi      a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
```

```
##  6 audi      a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
##  7 audi      a4      3.1  2008     6 auto~ f        18    27 p     comp~    4.24
##  8 audi      a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi      a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 audi      a4 q~   2    2008     4 manu~ 4        20    28 p     comp~    4.47
## # ... with 133 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

2.8 slice() para linhas

```
slice(dados, 1:5)
```

```
## # A tibble: 5 x 15
##   manufact~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##   <fct>      <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
## 1 audi       a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
## 2 audi       a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
## 3 audi       a4      2    2008     4 manu~ f        20    31 p     comp~    4.47
## 4 audi       a4      2    2008     4 auto~ f        21    30 p     comp~    4.58
## 5 audi       a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
## # ... with 3 more variables: `soma de variáveis` <dbl>, car <chr>,
## #   `cyl / trans` <chr>, and abbreviated variable names 1: manufacturer,
## #   2: sqrt_cty
```

```
# dados[1:5,]

slice(dados, 20:30)
```

```
## # A tibble: 11 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 chevrolet c150~   5.3  2008     8 auto~ r        11    15 e     suv      3.32
##  2 chevrolet c150~   5.3  2008     8 auto~ r        14    20 r     suv      3.74
##  3 chevrolet c150~   5.7  1999     8 auto~ r        13    17 r     suv      3.61
##  4 chevrolet c150~   6    2008     8 auto~ r        12    17 r     suv      3.46
##  5 chevrolet corv~   5.7  1999     8 manu~ r        16    26 p     2sea~    4
##  6 chevrolet corv~   5.7  1999     8 auto~ r        15    23 p     2sea~    3.87
##  7 chevrolet corv~   6.2  2008     8 manu~ r        16    26 p     2sea~    4
##  8 chevrolet corv~   6.2  2008     8 auto~ r        15    25 p     2sea~    3.87
##  9 chevrolet corv~   7    2008     8 manu~ r        15    24 p     2sea~    3.87
## 10 chevrolet k150~   5.3  2008     8 auto~ 4        14    19 r     suv      3.74
```

```
## 11 chevrolet k150~   5.3  2008     8 auto~ 4        11    14 e     suv      3.32
## # ... with 3 more variables: `soma de variáveis` <dbl>, car <chr>,
## #   `cyl / trans` <chr>, and abbreviated variable names 1: manufacturer,
## #   2: sqrt_cty
```

```
# dados[20:30,]
```

2.9 arrange() para linhas

```
# ordenar "displ" de menor a maior
arrange(dados, displ)
```

```
## # A tibble: 234 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 honda     civic   1.6  1999     4 manu~ f        28    33 r     subc~    5.29
##  2 honda     civic   1.6  1999     4 auto~ f        24    32 r     subc~    4.90
##  3 honda     civic   1.6  1999     4 manu~ f        25    32 r     subc~    5
##  4 honda     civic   1.6  1999     4 manu~ f        23    29 p     subc~    4.80
##  5 honda     civic   1.6  1999     4 auto~ f        24    32 r     subc~    4.90
##  6 audi      a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
##  7 audi      a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  8 audi      a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi      a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 honda     civic   1.8  2008     4 manu~ f        26    34 r     subc~    5.10
## # ... with 224 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

```
arrange(dados, displ) %>%
  print(n=20)
```

```
## # A tibble: 234 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 honda     civic   1.6  1999     4 manu~ f        28    33 r     subc~    5.29
##  2 honda     civic   1.6  1999     4 auto~ f        24    32 r     subc~    4.90
##  3 honda     civic   1.6  1999     4 manu~ f        25    32 r     subc~    5
##  4 honda     civic   1.6  1999     4 manu~ f        23    29 p     subc~    4.80
##  5 honda     civic   1.6  1999     4 auto~ f        24    32 r     subc~    4.90
##  6 audi      a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
```

```
##  7 audi      a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  8 audi      a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi      a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 honda     civic   1.8  2008     4 manu~ f        26    34 r     subc~    5.10
## 11 honda     civic   1.8  2008     4 auto~ f        25    36 r     subc~    5
## 12 honda     civic   1.8  2008     4 auto~ f        24    36 c     subc~    4.90
## 13 toyota    coro~   1.8  1999     4 auto~ f        24    30 r     comp~    4.90
## 14 toyota    coro~   1.8  1999     4 auto~ f        24    33 r     comp~    4.90
## 15 toyota    coro~   1.8  1999     4 manu~ f        26    35 r     comp~    5.10
## 16 toyota    coro~   1.8  2008     4 manu~ f        28    37 r     comp~    5.29
## 17 toyota    coro~   1.8  2008     4 auto~ f        26    35 r     comp~    5.10
## 18 volkswag~ pass~   1.8  1999     4 manu~ f        21    29 p     mids~    4.58
## 19 volkswag~ pass~   1.8  1999     4 auto~ f        18    29 p     mids~    4.24
## 20 volkswag~ jetta   1.9  1999     4 manu~ f        33    44 d     comp~    5.74
## # ... with 214 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

```
# ordenar "displ" de maior a  menor
arrange(dados, desc(displ))
```

```
## # A tibble: 234 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 chevrolet corv~   7    2008     8 manu~ r        15    24 p     2sea~    3.87
##  2 chevrolet k150~   6.5  1999     8 auto~ 4        14    17 d     suv      3.74
##  3 chevrolet corv~   6.2  2008     8 manu~ r        16    26 p     2sea~    4
##  4 chevrolet corv~   6.2  2008     8 auto~ r        15    25 p     2sea~    3.87
##  5 jeep      gran~   6.1  2008     8 auto~ 4        11    14 p     suv      3.32
##  6 chevrolet c150~   6    2008     8 auto~ r        12    17 r     suv      3.46
##  7 dodge     dura~   5.9  1999     8 auto~ 4        11    15 r     suv      3.32
##  8 dodge     ram ~   5.9  1999     8 auto~ 4        11    15 r     pick~    3.32
##  9 chevrolet c150~   5.7  1999     8 auto~ r        13    17 r     suv      3.61
## 10 chevrolet corv~   5.7  1999     8 manu~ r        16    26 p     2sea~    4
## # ... with 224 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

```
arrange(dados, desc(displ)) %>%
  print(n=20)
```

```
## # A tibble: 234 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
```

```
##  1 chevrolet corv~   7    2008     8 manu~ r        15    24 p     2sea~    3.87
##  2 chevrolet k150~   6.5  1999     8 auto~ 4        14    17 d     suv      3.74
##  3 chevrolet corv~   6.2  2008     8 manu~ r        16    26 p     2sea~    4
##  4 chevrolet corv~   6.2  2008     8 auto~ r        15    25 p     2sea~    3.87
##  5 jeep      gran~   6.1  2008     8 auto~ 4        11    14 p     suv      3.32
##  6 chevrolet c150~   6    2008     8 auto~ r        12    17 r     suv      3.46
##  7 dodge     dura~   5.9  1999     8 auto~ 4        11    15 r     suv      3.32
##  8 dodge     ram ~   5.9  1999     8 auto~ 4        11    15 r     pick~    3.32
##  9 chevrolet c150~   5.7  1999     8 auto~ r        13    17 r     suv      3.61
## 10 chevrolet corv~   5.7  1999     8 manu~ r        16    26 p     2sea~    4
## 11 chevrolet corv~   5.7  1999     8 auto~ r        15    23 p     2sea~    3.87
## 12 chevrolet k150~   5.7  1999     8 auto~ 4        11    15 r     suv      3.32
## 13 dodge     dura~   5.7  2008     8 auto~ 4        13    18 r     suv      3.61
## 14 dodge     ram ~   5.7  2008     8 auto~ 4        13    17 r     pick~    3.61
## 15 jeep      gran~   5.7  2008     8 auto~ 4        13    18 r     suv      3.61
## 16 toyota    land~   5.7  2008     8 auto~ 4        13    18 r     suv      3.61
## 17 nissan    path~   5.6  2008     8 auto~ 4        12    18 p     suv      3.46
## 18 ford      expe~   5.4  1999     8 auto~ r        11    17 r     suv      3.32
## 19 ford      expe~   5.4  2008     8 auto~ r        12    18 r     suv      3.46
## 20 ford      f150~   5.4  1999     8 auto~ 4        11    15 r     pick~    3.32
## # ... with 214 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

```
select(dados, displ, cty) %>%
  arrange(displ, cty) %>%
  print(n = 20)
```

```
## # A tibble: 234 x 2
##    displ   cty
##    <dbl> <int>
##  1   1.6    23
##  2   1.6    24
##  3   1.6    24
##  4   1.6    25
##  5   1.6    28
##  6   1.8    16
##  7   1.8    18
##  8   1.8    18
##  9   1.8    18
## 10   1.8    21
## 11   1.8    21
## 12   1.8    24
## 13   1.8    24
## 14   1.8    24
```

```
## 15   1.8    25
## 16   1.8    26
## 17   1.8    26
## 18   1.8    26
## 19   1.8    28
## 20   1.9    29
## # ... with 214 more rows
```

```
select(dados, displ, cty) %>%
  arrange(displ, desc(cty)) %>%
  print(n = 20)
```

```
## # A tibble: 234 x 2
##    displ   cty
##    <dbl> <int>
##  1   1.6    28
##  2   1.6    25
##  3   1.6    24
##  4   1.6    24
##  5   1.6    23
##  6   1.8    28
##  7   1.8    26
##  8   1.8    26
##  9   1.8    26
## 10   1.8    25
## 11   1.8    24
## 12   1.8    24
## 13   1.8    24
## 14   1.8    21
## 15   1.8    21
## 16   1.8    18
## 17   1.8    18
## 18   1.8    18
## 19   1.8    16
## 20   1.9    35
## # ... with 214 more rows
```

2.10 distinct() para linhas

```
dados_exemplo <- data.frame(id = 1:3,
                            name = c("John", "Max", "Julia"))
dados_exemplo
```

```
##   id  name
## 1  1  John
## 2  2   Max
## 3  3 Julia
```

```
# bind_rows == rbind()
dados_exemplo<- bind_rows(dados_exemplo, slice(dados_exemplo, 2))
dados_exemplo
```

```
##   id  name
## 1  1  John
## 2  2   Max
## 3  3 Julia
## 4  2   Max
```

```
distinct(dados_exemplo)
```

```
##   id  name
## 1  1  John
## 2  2   Max
## 3  3 Julia
```

```
dados_exemplo2 <- data.frame(id = c(1,1,2),
                             name = c("John", "Max", "Julia"))
dados_exemplo2
```

```
##   id  name
## 1  1  John
## 2  1   Max
## 3  2 Julia
```

```
distinct(dados_exemplo2)
```

```
##   id  name
## 1  1  John
## 2  1   Max
## 3  2 Julia
```

```
dados_duplicados <- select(dados, manufacturer, model)
dados_duplicados
```

```
## # A tibble: 234 x 2
##    manufacturer model
##    <fct>        <fct>
##  1 audi         a4
##  2 audi         a4
##  3 audi         a4
##  4 audi         a4
##  5 audi         a4
##  6 audi         a4
##  7 audi         a4
##  8 audi         a4 quattro
##  9 audi         a4 quattro
## 10 audi         a4 quattro
## # ... with 224 more rows
```

```
dados_nao_duplicados <- distinct(dados_duplicados)
dados_nao_duplicados
```

```
## # A tibble: 38 x 2
##    manufacturer model
##    <fct>        <fct>
##  1 audi         a4
##  2 audi         a4 quattro
##  3 audi         a6 quattro
##  4 chevrolet    c1500 suburban 2wd
##  5 chevrolet    corvette
##  6 chevrolet    k1500 tahoe 4wd
##  7 chevrolet    malibu
##  8 dodge        caravan 2wd
##  9 dodge        dakota pickup 4wd
## 10 dodge        durango 4wd
## # ... with 28 more rows
```

2.11 summarise()

```
summarise(dados, `média hwy` = mean(hwy))
```

```
## # A tibble: 1 x 1
##   `média hwy`
##         <dbl>
## 1        23.4
```

```
summarise(dados,
          `num. de dados` = n(),
          `num. modelos` = n_distinct(model))
```

```
## # A tibble: 1 x 2
##   `num. de dados` `num. modelos`
##             <int>          <int>
## 1             234             38
```

```
# levels(dados$model)
summarise(dados,
          `mín. hwy` = min(hwy, na.rm = TRUE),
          `mín. cty` = min(cty, na.rm = TRUE),
          `máx. hwy` = max(hwy, na.rm = TRUE),
          `máx. cty` = max(cty, na.rm = TRUE))
```

```
## # A tibble: 1 x 4
##   `mín. hwy` `mín. cty` `máx. hwy` `máx. cty`
##        <int>      <int>      <int>      <int>
## 1         12          9         44         35
```

```
dados %>%
  summarise_at(c("hwy", "cty"), list(min, max), na.rm = TRUE)
```

```
## # A tibble: 1 x 4
##   hwy_fn1 cty_fn1 hwy_fn2 cty_fn2
##     <int>   <int>   <int>   <int>
## 1      12       9      44      35
```

```
dados %>%
  summarise_if(is.numeric, list(min, max), na.rm = TRUE)
```

```
## # A tibble: 1 x 14
##   displ_fn1 year_fn1 cyl_fn1 cty_fn1 hwy_fn1 sqrt_cty_~1 soma ~2 displ~3 year_~4
##       <dbl>    <int>   <int>   <int>   <int>       <dbl>   <dbl>   <dbl>   <int>
## 1       1.6     1999       4       9      12           3    10.5       7    2008
## # ... with 5 more variables: cyl_fn2 <int>, cty_fn2 <int>, hwy_fn2 <int>,
## #   sqrt_cty_fn2 <dbl>, `soma de variáveis_fn2` <dbl>, and abbreviated variable
## #   names 1: sqrt_cty_fn1, 2: `soma de variáveis_fn1`, 3: displ_fn2,
## #   4: year_fn2
```

```
dados %>%
  summarise_if(is.numeric, min, na.rm = TRUE)
```

```
## # A tibble: 1 x 7
##   displ  year   cyl   cty   hwy sqrt_cty `soma de variáveis`
##   <dbl> <int> <int> <int> <int>    <dbl>               <dbl>
## 1   1.6  1999     4     9    12        3                10.5
```

```
dados %>%
  summarise_if(is.numeric, max, na.rm = TRUE)
```

```
## # A tibble: 1 x 7
##   displ  year   cyl   cty   hwy sqrt_cty `soma de variáveis`
##   <dbl> <int> <int> <int> <int>    <dbl>               <dbl>
## 1     7  2008     8    35    44     5.92                39.5
```

```
Tiago<- function(dados){
  sd(dados)/mean(dados)
}

dados %>%
  summarise_if(is.numeric, Tiago)
```

```
## # A tibble: 1 x 7
##   displ    year   cyl   cty   hwy sqrt_cty `soma de variáveis`
##   <dbl>   <dbl> <dbl> <dbl> <dbl>    <dbl>               <dbl>
## 1 0.372 0.00225 0.274 0.252 0.254    0.125               0.251
```

2.12 group_by()

```
group_by(dados, manufacturer)
```

```
## # A tibble: 234 x 15
## # Groups:   manufacturer [15]
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 audi      a4      1.8  1999     4 auto~ f        18    29 p     comp~    4.24
##  2 audi      a4      1.8  1999     4 manu~ f        21    29 p     comp~    4.58
##  3 audi      a4      2    2008     4 manu~ f        20    31 p     comp~    4.47
##  4 audi      a4      2    2008     4 auto~ f        21    30 p     comp~    4.58
```

```
##  5 audi      a4      2.8  1999     6 auto~ f        16    26 p     comp~    4
##  6 audi      a4      2.8  1999     6 manu~ f        18    26 p     comp~    4.24
##  7 audi      a4      3.1  2008     6 auto~ f        18    27 p     comp~    4.24
##  8 audi      a4 q~   1.8  1999     4 manu~ 4        18    26 p     comp~    4.24
##  9 audi      a4 q~   1.8  1999     4 auto~ 4        16    25 p     comp~    4
## 10 audi      a4 q~   2    2008     4 manu~ 4        20    28 p     comp~    4.47
## # ... with 224 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

```
dados %>%
  group_by(manufacturer) %>%
  summarise(`num. carros` = n())
```

```
## # A tibble: 15 x 2
##    manufacturer `num. carros`
##    <fct>                <int>
##  1 audi                    18
##  2 chevrolet               19
##  3 dodge                   37
##  4 ford                    25
##  5 honda                    9
##  6 hyundai                 14
##  7 jeep                     8
##  8 land rover               4
##  9 lincoln                  3
## 10 mercury                  4
## 11 nissan                  13
## 12 pontiac                  5
## 13 subaru                  14
## 14 toyota                  34
## 15 volkswagen              27
```

```
dados %>%
  group_by(model) %>%
  summarise(`média hwy` = mean(hwy),
            `min. hwy` = min(hwy),
            `max. hwy` = max(hwy))
```

```
## # A tibble: 38 x 4
##    model              `média hwy` `min. hwy` `max. hwy`
##    <fct>                    <dbl>      <int>      <int>
##  1 4runner 4wd               18.8         17         20
##  2 a4                        28.3         26         31
```

```
##  3 a4 quattro               25.8         25         28
##  4 a6 quattro               24           23         25
##  5 altima                   28.7         26         32
##  6 c1500 suburban 2wd       17.8         15         20
##  7 camry                    28.3         26         31
##  8 camry solara             28.1         26         31
##  9 caravan 2wd              22.4         17         24
## 10 civic                    32.6         29         36
## # ... with 28 more rows
```

2.13 count()

```
count(dados)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   234
```

```
dados %>%
  group_by(manufacturer) %>%
  count()
```

```
## # A tibble: 15 x 2
## # Groups:   manufacturer [15]
##    manufacturer     n
##    <fct>        <int>
##  1 audi            18
##  2 chevrolet       19
##  3 dodge           37
##  4 ford            25
##  5 honda            9
##  6 hyundai         14
##  7 jeep             8
##  8 land rover       4
##  9 lincoln          3
## 10 mercury          4
## 11 nissan          13
## 12 pontiac          5
## 13 subaru          14
## 14 toyota          34
## 15 volkswagen      27
```

```
# Equivalente com o código anterior
dados %>%
  group_by(manufacturer) %>%
  summarise(cars = n())
```

```
## # A tibble: 15 x 2
##    manufacturer  cars
##    <fct>        <int>
##  1 audi            18
##  2 chevrolet       19
##  3 dodge           37
##  4 ford            25
##  5 honda            9
##  6 hyundai         14
##  7 jeep             8
##  8 land rover       4
##  9 lincoln          3
## 10 mercury          4
## 11 nissan          13
## 12 pontiac          5
## 13 subaru          14
## 14 toyota          34
## 15 volkswagen      27
```

2.14 sample_n()

```
set.seed(567)
sample_n(dados, size = 10, replace = F)
```

```
## # A tibble: 10 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 mercury   moun~   5    1999     8 auto~ 4        13    17 r     suv      3.61
##  2 chevrolet corv~   7    2008     8 manu~ r        15    24 p     2sea~    3.87
##  3 dodge     ram ~   4.7  2008     8 manu~ 4        12    16 r     pick~    3.46
##  4 toyota    land~   4.7  1999     8 auto~ 4        11    15 r     suv      3.32
##  5 volkswag~ jetta   2    1999     4 auto~ f        19    26 r     comp~    4.36
##  6 dodge     cara~   3.8  1999     6 auto~ f        15    21 r     mini~    3.87
##  7 honda     civic   1.8  2008     4 auto~ f        25    36 r     subc~    5
##  8 ford      must~   4.6  1999     8 auto~ r        15    21 r     subc~    3.87
##  9 chevrolet c150~   5.3  2008     8 auto~ r        14    20 r     suv      3.74
## 10 ford      expe~   5.4  1999     8 auto~ r        11    17 r     suv      3.32
```

```
## # ... with 3 more variables: `soma de variáveis` <dbl>, car <chr>,
## #   `cyl / trans` <chr>, and abbreviated variable names 1: manufacturer,
## #   2: sqrt_cty
```

```
sample_n(dados, size = 10, replace = T)
```

```
## # A tibble: 10 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 chevrolet c150~   5.3  2008     8 auto~ r        11    15 e     suv      3.32
##  2 volkswag~ gti     2    2008     4 auto~ f        22    29 p     comp~    4.69
##  3 dodge     dako~   4.7  2008     8 auto~ 4        14    19 r     pick~    3.74
##  4 ford      expl~   4.6  2008     8 auto~ 4        13    19 r     suv      3.61
##  5 dodge     cara~   3.8  2008     6 auto~ f        16    23 r     mini~    4
##  6 chevrolet k150~   5.3  2008     8 auto~ 4        14    19 r     suv      3.74
##  7 dodge     dura~   5.2  1999     8 auto~ 4        11    16 r     suv      3.32
##  8 toyota    camry   2.4  2008     4 manu~ f        21    31 r     mids~    4.58
##  9 toyota    camry   3    1999     6 manu~ f        18    26 r     mids~    4.24
## 10 subaru    impr~   2.2  1999     4 auto~ 4        21    26 r     subc~    4.58
## # ... with 3 more variables: `soma de variáveis` <dbl>, car <chr>,
## #   `cyl / trans` <chr>, and abbreviated variable names 1: manufacturer,
## #   2: sqrt_cty
```

2.15 sample_frac()

```
sample_frac(dados, size = 0.1, replace = F)
```

```
## # A tibble: 23 x 15
##    manufac~1 model displ  year   cyl trans drv     cty   hwy fl    class sqrt_~2
##    <fct>     <fct> <dbl> <int> <int> <fct> <fct> <int> <int> <fct> <fct>   <dbl>
##  1 toyota    coro~   1.8  2008     4 manu~ f        28    37 r     comp~    5.29
##  2 lincoln   navi~   5.4  1999     8 auto~ r        11    17 r     suv      3.32
##  3 honda     civic   1.6  1999     4 auto~ f        24    32 r     subc~    4.90
##  4 audi      a6 q~   2.8  1999     6 auto~ 4        15    24 p     mids~    3.87
##  5 nissan    path~   4    2008     6 auto~ 4        14    20 p     suv      3.74
##  6 toyota    camry   3.5  2008     6 auto~ f        19    28 r     mids~    4.36
##  7 subaru    impr~   2.5  2008     4 auto~ 4        20    25 p     comp~    4.47
##  8 toyota    toyo~   3.4  1999     6 auto~ 4        15    19 r     pick~    3.87
##  9 audi      a4 q~   3.1  2008     6 manu~ 4        15    25 p     comp~    3.87
## 10 toyota    coro~   1.8  1999     4 manu~ f        26    35 r     comp~    5.10
## # ... with 13 more rows, 3 more variables: `soma de variáveis` <dbl>,
## #   car <chr>, `cyl / trans` <chr>, and abbreviated variable names
## #   1: manufacturer, 2: sqrt_cty
```

Capítulo 3

ggplot2 (60 minutos)

Alguns links

The R Graph Gallery

120 registered extensions available to explore

link 1: patchwork

link 2: patchwork

3.1 Lista de funções do pacote ggplot2

```
ls("package:ggplot2")
```

```
##   [1] "%+%"                       "%+replace%"
##   [3] "aes"                       "aes_"
##   [5] "aes_all"                   "aes_auto"
##   [7] "aes_q"                     "aes_string"
##   [9] "after_scale"               "after_stat"
##  [11] "alpha"                     "annotate"
##  [13] "annotation_custom"         "annotation_logticks"
##  [15] "annotation_map"            "annotation_raster"
##  [17] "arrow"                     "as_label"
##  [19] "as_labeller"               "autolayer"
##  [21] "autoplot"                  "AxisSecondary"
##  [23] "benchplot"                 "binned_scale"
##  [25] "borders"                   "calc_element"
##  [27] "combine_vars"              "continuous_scale"
```

```
##  [29] "Coord"                      "coord_cartesian"
##  [31] "coord_equal"                "coord_fixed"
##  [33] "coord_flip"                 "coord_map"
##  [35] "coord_munch"                "coord_polar"
##  [37] "coord_quickmap"             "coord_sf"
##  [39] "coord_trans"                "CoordCartesian"
##  [41] "CoordFixed"                 "CoordFlip"
##  [43] "CoordMap"                   "CoordPolar"
##  [45] "CoordQuickmap"              "CoordSf"
##  [47] "CoordTrans"                 "cut_interval"
##  [49] "cut_number"                 "cut_width"
##  [51] "derive"                     "diamonds"
##  [53] "discrete_scale"             "draw_key_abline"
##  [55] "draw_key_blank"             "draw_key_boxplot"
##  [57] "draw_key_crossbar"          "draw_key_dotplot"
##  [59] "draw_key_label"             "draw_key_linerange"
##  [61] "draw_key_path"              "draw_key_point"
##  [63] "draw_key_pointrange"        "draw_key_polygon"
##  [65] "draw_key_rect"              "draw_key_smooth"
##  [67] "draw_key_text"              "draw_key_timeseries"
##  [69] "draw_key_vline"             "draw_key_vpath"
##  [71] "dup_axis"                   "economics"
##  [73] "economics_long"             "el_def"
##  [75] "element_blank"              "element_grob"
##  [77] "element_line"               "element_rect"
##  [79] "element_render"             "element_text"
##  [81] "enexpr"                     "enexprs"
##  [83] "enquo"                      "enquos"
##  [85] "ensym"                      "ensyms"
##  [87] "expand_limits"              "expand_scale"
##  [89] "expansion"                  "expr"
##  [91] "Facet"                      "facet_grid"
##  [93] "facet_null"                 "facet_wrap"
##  [95] "FacetGrid"                  "FacetNull"
##  [97] "FacetWrap"                  "faithfuld"
##  [99] "find_panel"                 "flip_data"
## [101] "flipped_names"              "fortify"
## [103] "Geom"                       "geom_abline"
## [105] "geom_area"                  "geom_bar"
## [107] "geom_bin_2d"                "geom_bin2d"
## [109] "geom_blank"                 "geom_boxplot"
## [111] "geom_col"                   "geom_contour"
## [113] "geom_contour_filled"        "geom_count"
## [115] "geom_crossbar"              "geom_curve"
## [117] "geom_density"               "geom_density_2d"
## [119] "geom_density_2d_filled"     "geom_density2d"
```

```
## [121] "geom_density2d_filled"      "geom_dotplot"
## [123] "geom_errorbar"              "geom_errorbarh"
## [125] "geom_freqpoly"              "geom_function"
## [127] "geom_hex"                   "geom_histogram"
## [129] "geom_hline"                 "geom_jitter"
## [131] "geom_label"                 "geom_line"
## [133] "geom_linerange"             "geom_map"
## [135] "geom_path"                  "geom_point"
## [137] "geom_pointrange"            "geom_polygon"
## [139] "geom_qq"                    "geom_qq_line"
## [141] "geom_quantile"              "geom_raster"
## [143] "geom_rect"                  "geom_ribbon"
## [145] "geom_rug"                   "geom_segment"
## [147] "geom_sf"                    "geom_sf_label"
## [149] "geom_sf_text"               "geom_smooth"
## [151] "geom_spoke"                 "geom_step"
## [153] "geom_text"                  "geom_tile"
## [155] "geom_violin"                "geom_vline"
## [157] "GeomAbline"                 "GeomAnnotationMap"
## [159] "GeomArea"                   "GeomBar"
## [161] "GeomBlank"                  "GeomBoxplot"
## [163] "GeomCol"                    "GeomContour"
## [165] "GeomContourFilled"          "GeomCrossbar"
## [167] "GeomCurve"                  "GeomCustomAnn"
## [169] "GeomDensity"                "GeomDensity2d"
## [171] "GeomDensity2dFilled"        "GeomDotplot"
## [173] "GeomErrorbar"               "GeomErrorbarh"
## [175] "GeomFunction"               "GeomHex"
## [177] "GeomHline"                  "GeomLabel"
## [179] "GeomLine"                   "GeomLinerange"
## [181] "GeomLogticks"               "GeomMap"
## [183] "GeomPath"                   "GeomPoint"
## [185] "GeomPointrange"             "GeomPolygon"
## [187] "GeomQuantile"               "GeomRaster"
## [189] "GeomRasterAnn"              "GeomRect"
## [191] "GeomRibbon"                 "GeomRug"
## [193] "GeomSegment"                "GeomSf"
## [195] "GeomSmooth"                 "GeomSpoke"
## [197] "GeomStep"                   "GeomText"
## [199] "GeomTile"                   "GeomViolin"
## [201] "GeomVline"                  "get_alt_text"
## [203] "get_element_tree"           "gg_dep"
## [205] "ggplot"                     "ggplot_add"
## [207] "ggplot_build"               "ggplot_gtable"
## [209] "ggplotGrob"                 "ggproto"
## [211] "ggproto_parent"             "ggsave"
```

```
## [213] "ggtitle"                    "guide_axis"
## [215] "guide_bins"                 "guide_colorbar"
## [217] "guide_colorsteps"           "guide_colourbar"
## [219] "guide_coloursteps"          "guide_gengrob"
## [221] "guide_geom"                 "guide_legend"
## [223] "guide_merge"                "guide_none"
## [225] "guide_train"                "guide_transform"
## [227] "guides"                     "has_flipped_aes"
## [229] "is.Coord"                   "is.facet"
## [231] "is.ggplot"                  "is.ggproto"
## [233] "is.theme"                   "label_both"
## [235] "label_bquote"               "label_context"
## [237] "label_parsed"               "label_value"
## [239] "label_wrap_gen"             "labeller"
## [241] "labs"                       "last_plot"
## [243] "layer"                      "layer_data"
## [245] "layer_grob"                 "layer_scales"
## [247] "layer_sf"                   "Layout"
## [249] "lims"                       "luv_colours"
## [251] "map_data"                   "margin"
## [253] "max_height"                 "max_width"
## [255] "mean_cl_boot"               "mean_cl_normal"
## [257] "mean_sdl"                   "mean_se"
## [259] "median_hilow"               "merge_element"
## [261] "midwest"                    "mpg"
## [263] "msleep"                     "panel_cols"
## [265] "panel_rows"                 "Position"
## [267] "position_dodge"             "position_dodge2"
## [269] "position_fill"              "position_identity"
## [271] "position_jitter"            "position_jitterdodge"
## [273] "position_nudge"             "position_stack"
## [275] "PositionDodge"              "PositionDodge2"
## [277] "PositionFill"               "PositionIdentity"
## [279] "PositionJitter"             "PositionJitterdodge"
## [281] "PositionNudge"              "PositionStack"
## [283] "presidential"               "qplot"
## [285] "quickplot"                  "quo"
## [287] "quo_name"                   "quos"
## [289] "register_theme_elements"    "rel"
## [291] "remove_missing"             "render_axes"
## [293] "render_strips"              "reset_theme_settings"
## [295] "resolution"                 "Scale"
## [297] "scale_alpha"                "scale_alpha_binned"
## [299] "scale_alpha_continuous"     "scale_alpha_date"
## [301] "scale_alpha_datetime"       "scale_alpha_discrete"
## [303] "scale_alpha_identity"       "scale_alpha_manual"
```

```
## [305] "scale_alpha_ordinal"        "scale_color_binned"
## [307] "scale_color_brewer"         "scale_color_continuous"
## [309] "scale_color_date"           "scale_color_datetime"
## [311] "scale_color_discrete"       "scale_color_distiller"
## [313] "scale_color_fermenter"      "scale_color_gradient"
## [315] "scale_color_gradient2"      "scale_color_gradientn"
## [317] "scale_color_grey"           "scale_color_hue"
## [319] "scale_color_identity"       "scale_color_manual"
## [321] "scale_color_ordinal"        "scale_color_steps"
## [323] "scale_color_steps2"         "scale_color_stepsn"
## [325] "scale_color_viridis_b"      "scale_color_viridis_c"
## [327] "scale_color_viridis_d"      "scale_colour_binned"
## [329] "scale_colour_brewer"        "scale_colour_continuous"
## [331] "scale_colour_date"          "scale_colour_datetime"
## [333] "scale_colour_discrete"      "scale_colour_distiller"
## [335] "scale_colour_fermenter"     "scale_colour_gradient"
## [337] "scale_colour_gradient2"     "scale_colour_gradientn"
## [339] "scale_colour_grey"          "scale_colour_hue"
## [341] "scale_colour_identity"      "scale_colour_manual"
## [343] "scale_colour_ordinal"       "scale_colour_steps"
## [345] "scale_colour_steps2"        "scale_colour_stepsn"
## [347] "scale_colour_viridis_b"     "scale_colour_viridis_c"
## [349] "scale_colour_viridis_d"     "scale_continuous_identity"
## [351] "scale_discrete_identity"    "scale_discrete_manual"
## [353] "scale_fill_binned"          "scale_fill_brewer"
## [355] "scale_fill_continuous"      "scale_fill_date"
## [357] "scale_fill_datetime"        "scale_fill_discrete"
## [359] "scale_fill_distiller"       "scale_fill_fermenter"
## [361] "scale_fill_gradient"        "scale_fill_gradient2"
## [363] "scale_fill_gradientn"       "scale_fill_grey"
## [365] "scale_fill_hue"             "scale_fill_identity"
## [367] "scale_fill_manual"          "scale_fill_ordinal"
## [369] "scale_fill_steps"           "scale_fill_steps2"
## [371] "scale_fill_stepsn"          "scale_fill_viridis_b"
## [373] "scale_fill_viridis_c"       "scale_fill_viridis_d"
## [375] "scale_linetype"             "scale_linetype_binned"
## [377] "scale_linetype_continuous"  "scale_linetype_discrete"
## [379] "scale_linetype_identity"    "scale_linetype_manual"
## [381] "scale_linewidth"            "scale_linewidth_binned"
## [383] "scale_linewidth_continuous" "scale_linewidth_date"
## [385] "scale_linewidth_datetime"   "scale_linewidth_discrete"
## [387] "scale_linewidth_ordinal"    "scale_radius"
## [389] "scale_shape"                "scale_shape_binned"
## [391] "scale_shape_continuous"     "scale_shape_discrete"
## [393] "scale_shape_identity"       "scale_shape_manual"
## [395] "scale_shape_ordinal"        "scale_size"
```

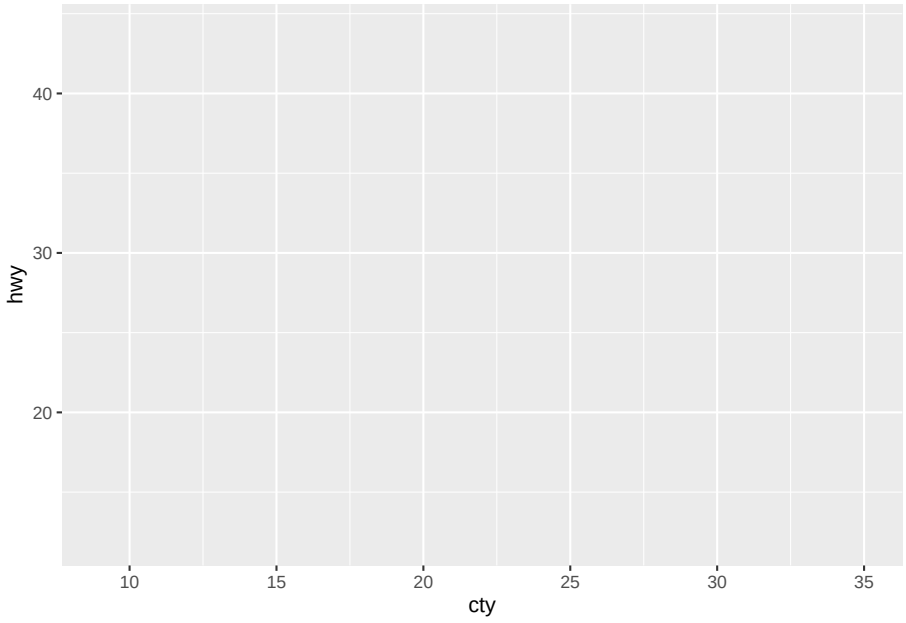
```
## [397] "scale_size_area"            "scale_size_binned"
## [399] "scale_size_binned_area"     "scale_size_continuous"
## [401] "scale_size_date"            "scale_size_datetime"
## [403] "scale_size_discrete"        "scale_size_identity"
## [405] "scale_size_manual"          "scale_size_ordinal"
## [407] "scale_type"                 "scale_x_binned"
## [409] "scale_x_continuous"         "scale_x_date"
## [411] "scale_x_datetime"           "scale_x_discrete"
## [413] "scale_x_log10"              "scale_x_reverse"
## [415] "scale_x_sqrt"               "scale_x_time"
## [417] "scale_y_binned"             "scale_y_continuous"
## [419] "scale_y_date"               "scale_y_datetime"
## [421] "scale_y_discrete"           "scale_y_log10"
## [423] "scale_y_reverse"            "scale_y_sqrt"
## [425] "scale_y_time"               "ScaleBinned"
## [427] "ScaleBinnedPosition"        "ScaleContinuous"
## [429] "ScaleContinuousDate"        "ScaleContinuousDatetime"
## [431] "ScaleContinuousIdentity"    "ScaleContinuousPosition"
## [433] "ScaleDiscrete"              "ScaleDiscreteIdentity"
## [435] "ScaleDiscretePosition"      "seals"
## [437] "sec_axis"                   "set_last_plot"
## [439] "sf_transform_xy"            "should_stop"
## [441] "stage"                      "standardise_aes_names"
## [443] "stat"                       "Stat"
## [445] "stat_align"                 "stat_bin"
## [447] "stat_bin_2d"                "stat_bin_hex"
## [449] "stat_bin2d"                 "stat_binhex"
## [451] "stat_boxplot"               "stat_contour"
## [453] "stat_contour_filled"        "stat_count"
## [455] "stat_density"               "stat_density_2d"
## [457] "stat_density_2d_filled"     "stat_density2d"
## [459] "stat_density2d_filled"      "stat_ecdf"
## [461] "stat_ellipse"               "stat_function"
## [463] "stat_identity"              "stat_qq"
## [465] "stat_qq_line"               "stat_quantile"
## [467] "stat_sf"                    "stat_sf_coordinates"
## [469] "stat_smooth"                "stat_spoke"
## [471] "stat_sum"                   "stat_summary"
## [473] "stat_summary_2d"            "stat_summary_bin"
## [475] "stat_summary_hex"           "stat_summary2d"
## [477] "stat_unique"                "stat_ydensity"
## [479] "StatAlign"                  "StatBin"
## [481] "StatBin2d"                  "StatBindot"
## [483] "StatBinhex"                 "StatBoxplot"
## [485] "StatContour"                "StatContourFilled"
## [487] "StatCount"                  "StatDensity"
```

```
## [489] "StatDensity2d"             "StatDensity2dFilled"
## [491] "StatEcdf"                  "StatEllipse"
## [493] "StatFunction"              "StatIdentity"
## [495] "StatQq"                    "StatQqLine"
## [497] "StatQuantile"              "StatSf"
## [499] "StatSfCoordinates"         "StatSmooth"
## [501] "StatSum"                   "StatSummary"
## [503] "StatSummary2d"             "StatSummaryBin"
## [505] "StatSummaryHex"            "StatUnique"
## [507] "StatYdensity"              "summarise_coord"
## [509] "summarise_layers"          "summarise_layout"
## [511] "sym"                       "syms"
## [513] "theme"                     "theme_bw"
## [515] "theme_classic"             "theme_dark"
## [517] "theme_get"                 "theme_gray"
## [519] "theme_grey"                "theme_light"
## [521] "theme_linedraw"            "theme_minimal"
## [523] "theme_replace"             "theme_set"
## [525] "theme_test"                "theme_update"
## [527] "theme_void"                "transform_position"
## [529] "txhousing"                 "unit"
## [531] "update_geom_defaults"      "update_labels"
## [533] "update_stat_defaults"      "vars"
## [535] "waiver"                    "wrap_dims"
## [537] "xlab"                      "xlim"
## [539] "ylab"                      "ylim"
## [541] "zeroGrob"
```

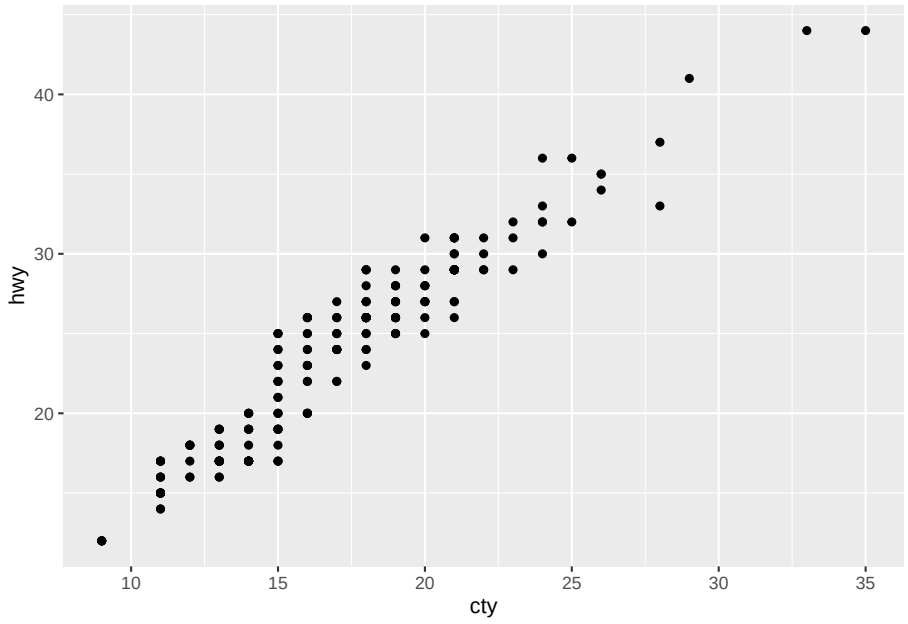
3.2 Primeiros passos usando geom_point

```
ggplot(dados)
```

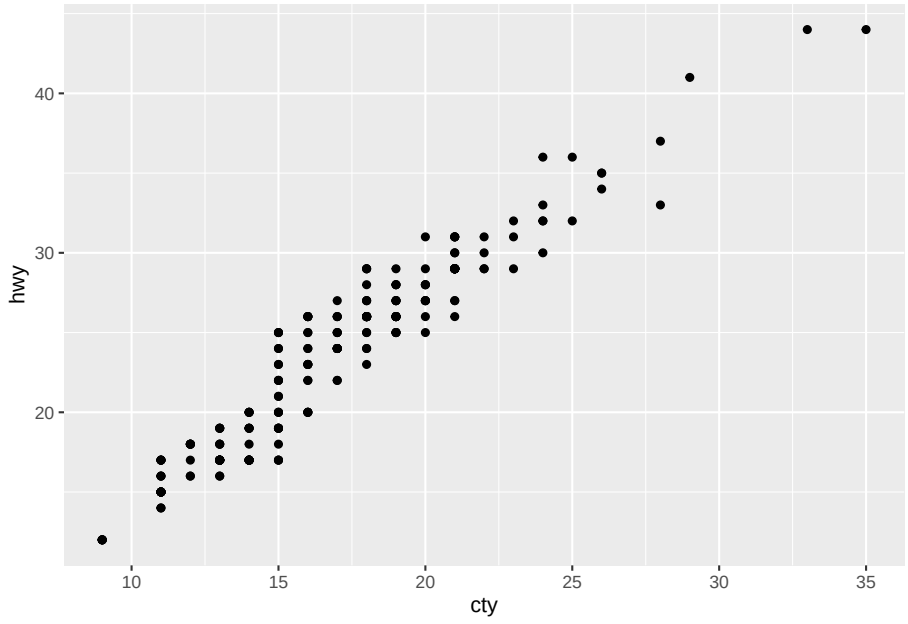
```
ggplot(dados, aes(x = cty, y = hwy))
```



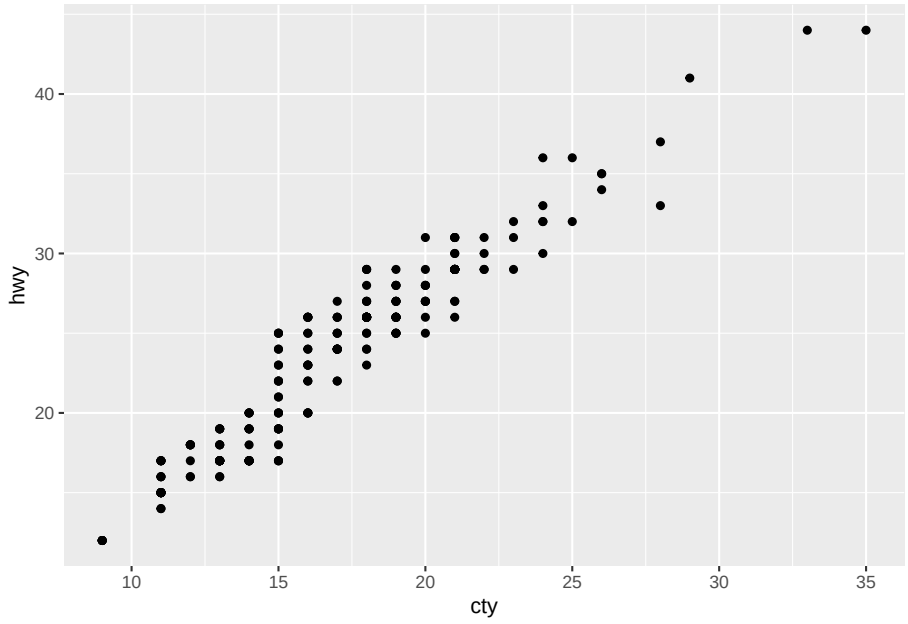
```
# Alternativas
ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point()
```



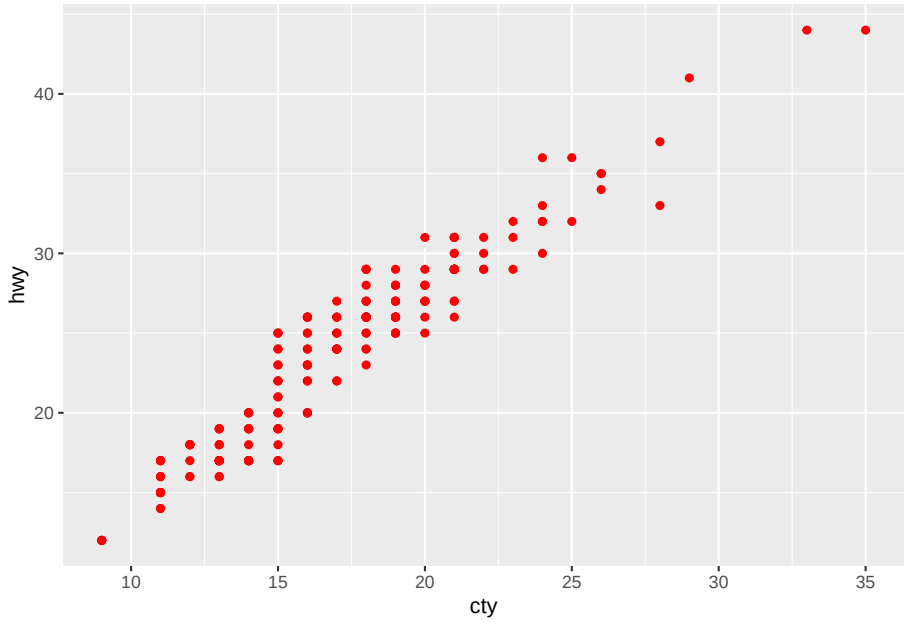

```
ggplot(dados) +
  geom_point(aes(x = cty, y = hwy))
```



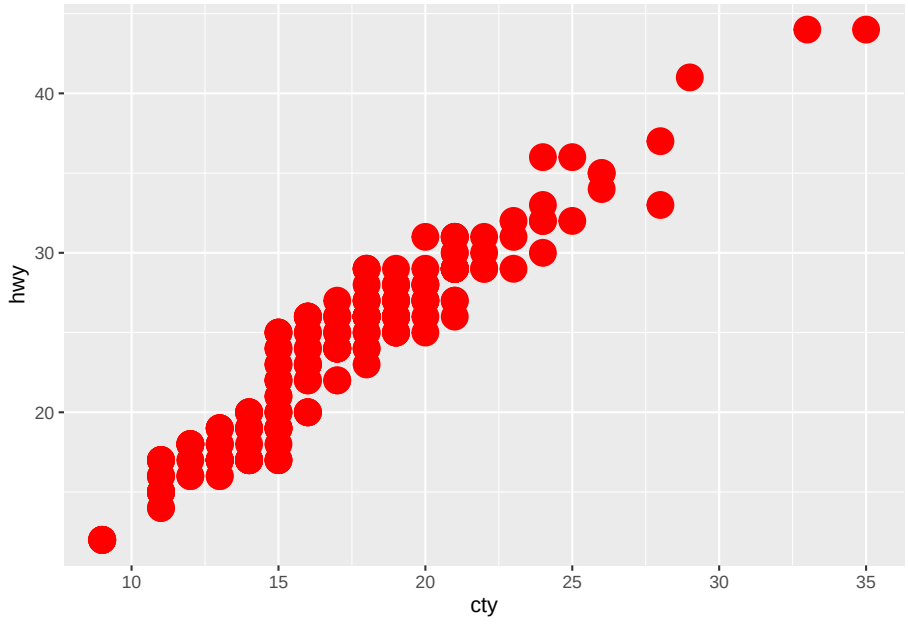

```
ggplot() +
  geom_point(data = dados, aes(x = cty, y = hwy))
```



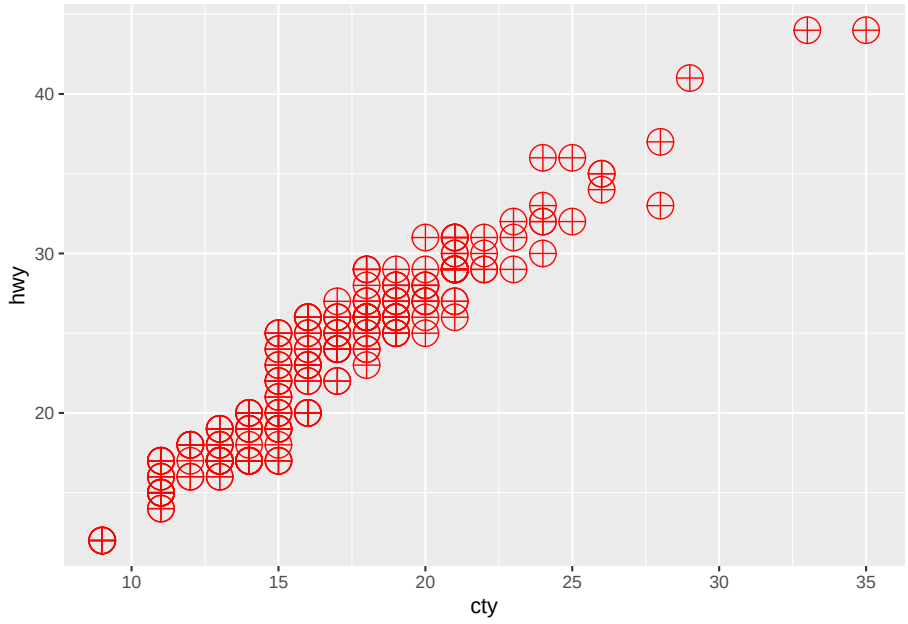
```
# Fim

ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point(colour = "red")
```



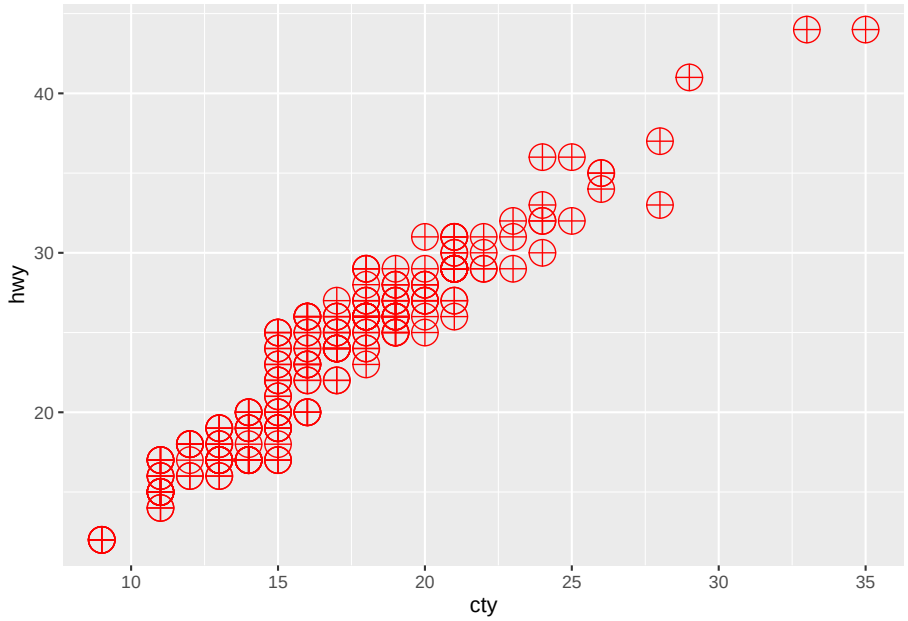

```
ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point(colour = "red", size = 6)
```



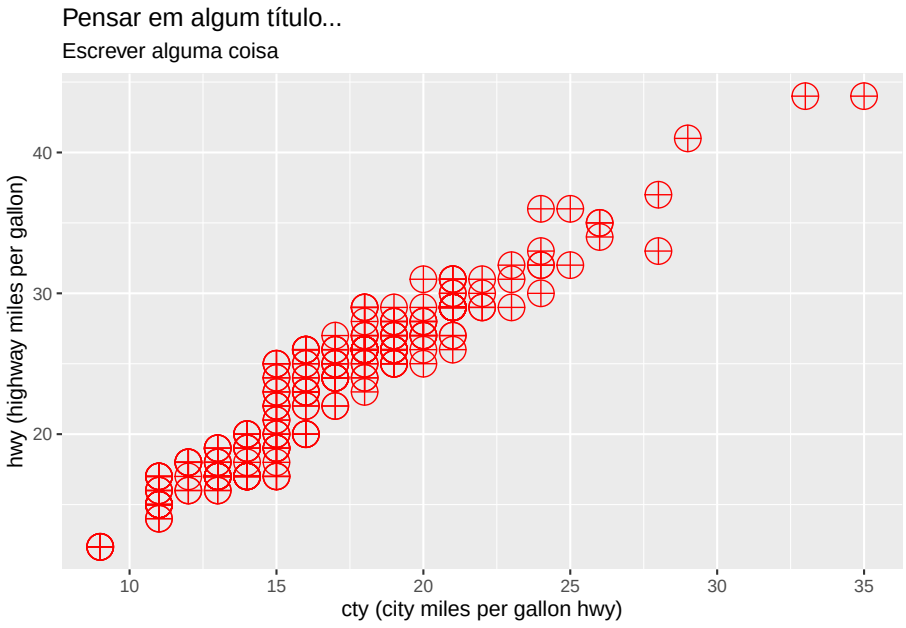

```
ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point(colour = "red", size = 6, shape = 10)
```



```
# Alternativa
ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point(colour = "red", size = 6, shape = "circle plus")
```




```
ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point(colour = "red", size = 6, shape = 10)+
  labs(x = "cty (city miles per gallon hwy)",
       y = "hwy (highway miles per gallon)",
       title = "Pensar em algum título...",
       subtitle = "Escrever alguma coisa")
```

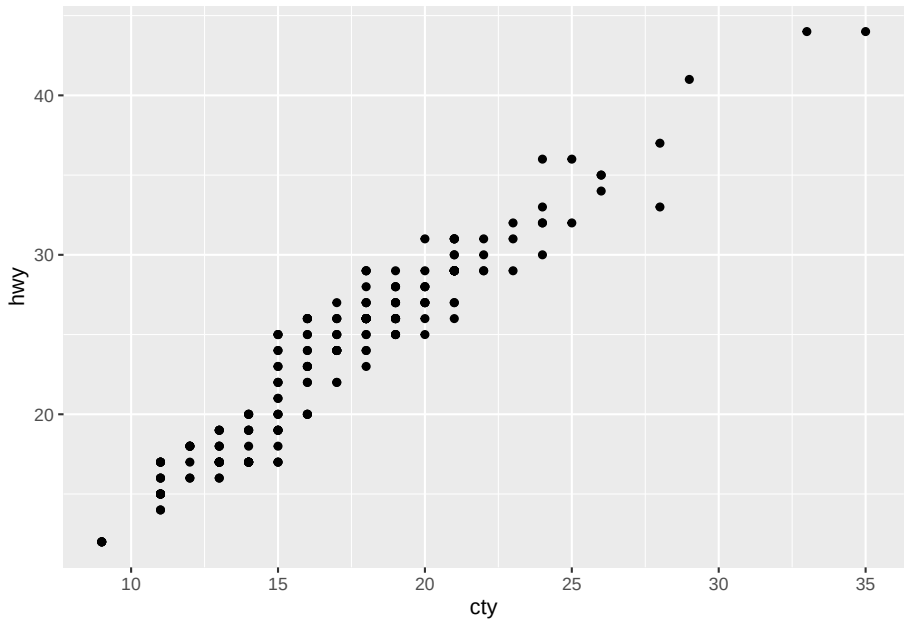


3.2.1 Mais detalhes sobre geom_point

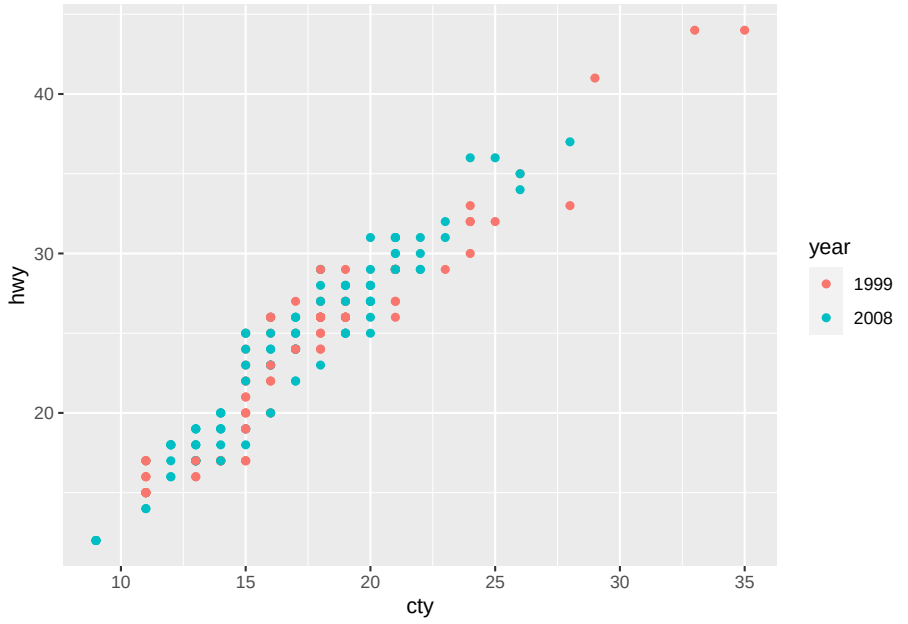
geom_point() understands the following aesthetics (required aesthetics are in bold):

- x
- y
- alpha
- colour
- fill
- group
- shape
- size
- stroke

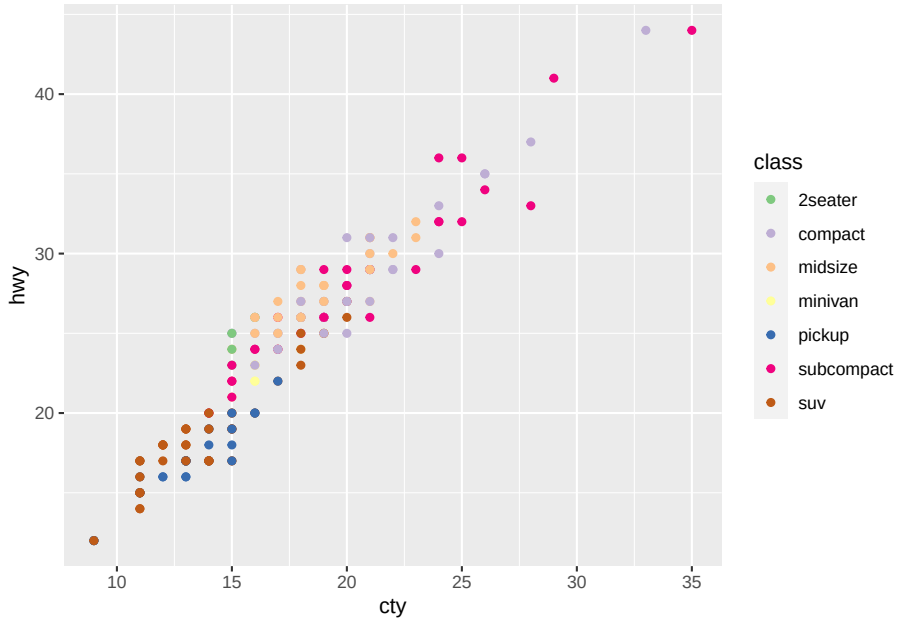
```
ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point()
```



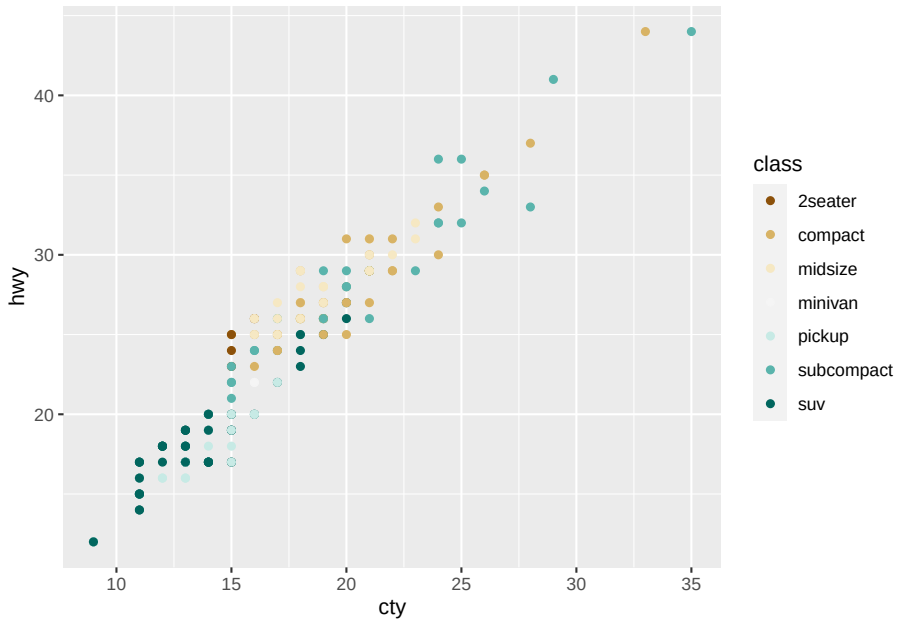

```
ggplot(dados, aes(x = cty, y = hwy, col = factor(year))) +
  geom_point() +
  labs(col = "year")
```



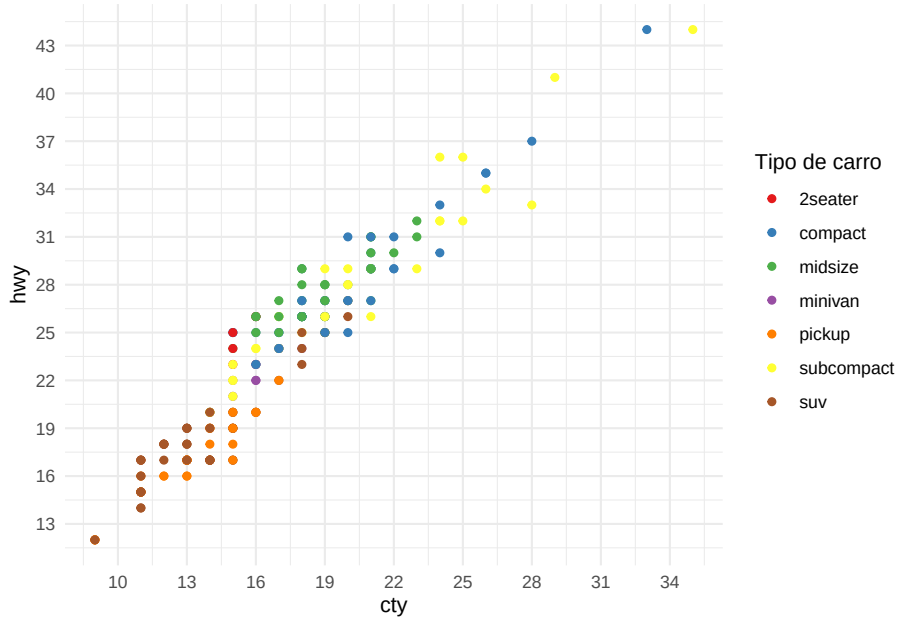
```
# Alternativa
ggplot(dados, aes(x = cty, y = hwy, col = factor(class))) +
  geom_point() +
  labs(col = "class")+
  scale_color_brewer(type = "qual")
```



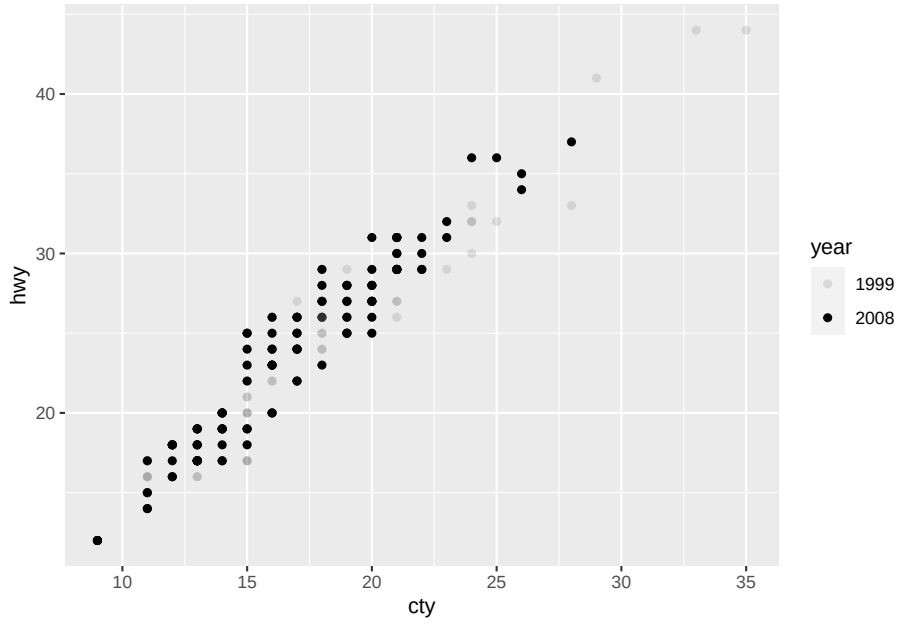

```
ggplot(dados, aes(x = cty, y = hwy, col = factor(class))) +
  geom_point() +
  labs(col = "class")+
  scale_color_brewer(type = "div")
```



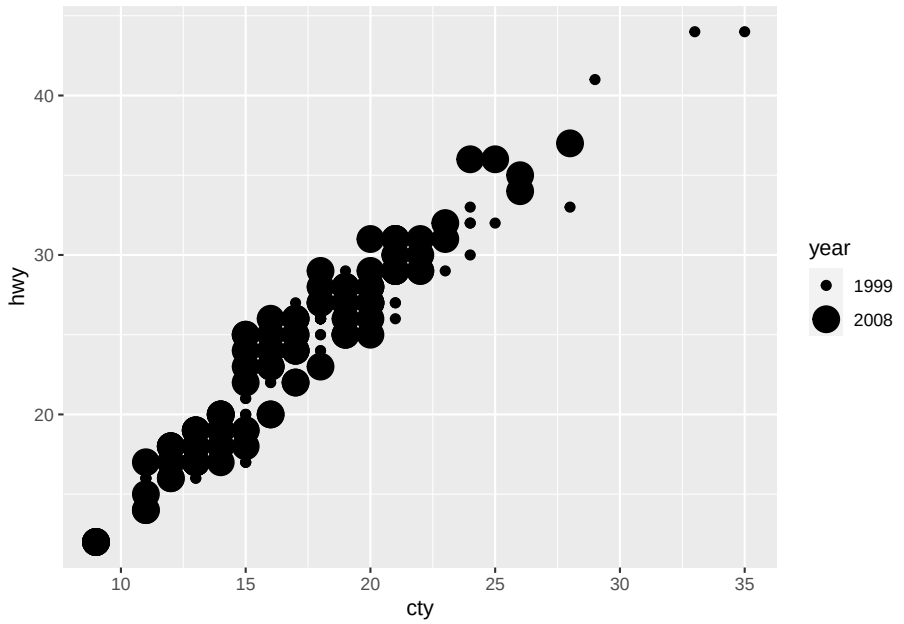

```
ggplot(dados, aes(x = cty, y = hwy, col = factor(class))) +
  geom_point() +
  labs(col = "class")+
  scale_color_brewer(palette = "Set1", name = "Tipo de carro")+
  scale_y_continuous(breaks = seq(10,60,3))+
  scale_x_continuous(breaks = seq(10,40,3))+
  theme_minimal()
```



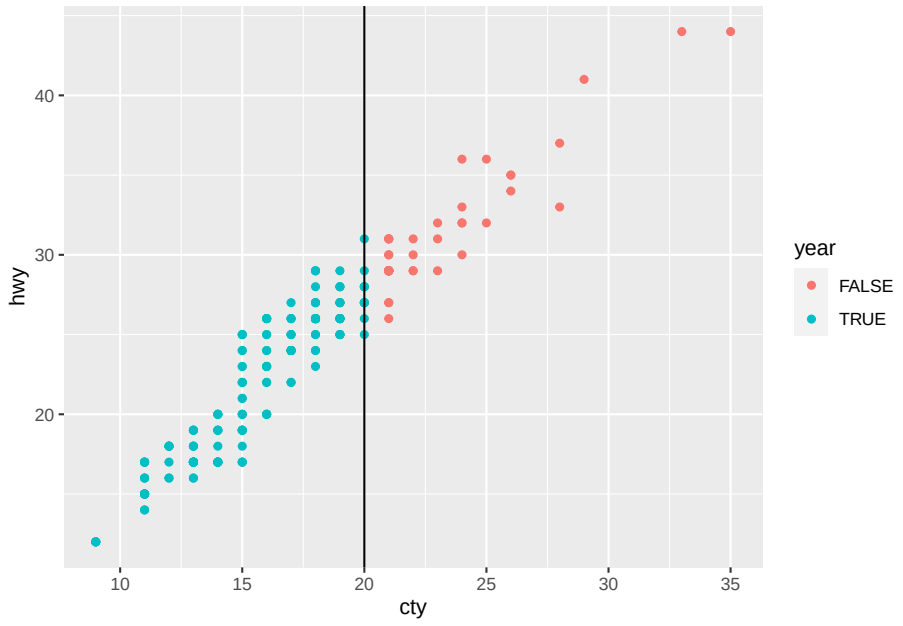

```
ggplot(dados, aes(x = cty, y = hwy, alpha = factor(year))) +
  geom_point() +
  labs(alpha = "year")
```



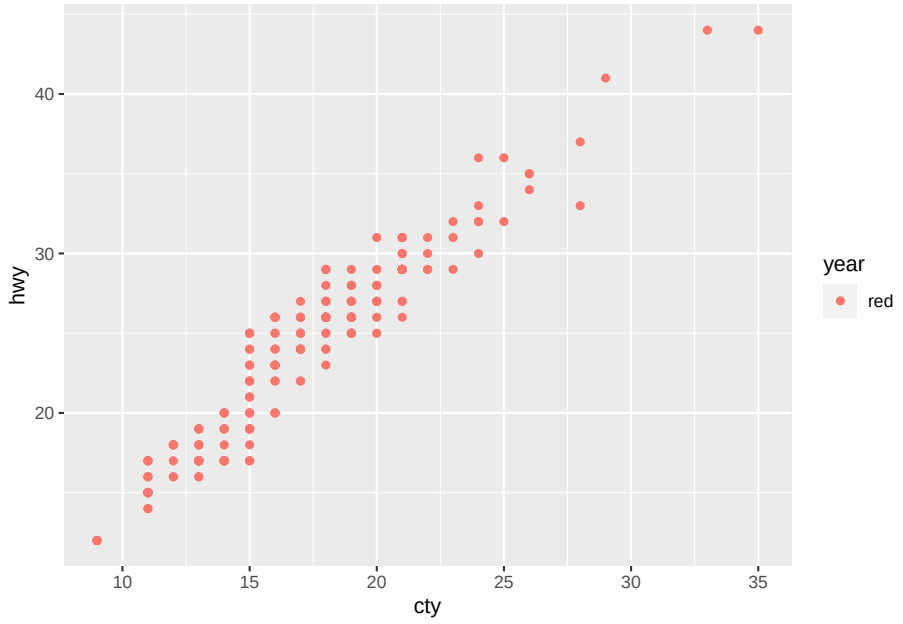
```
ggplot(dados, aes(x = cty, y = hwy, size = factor(year))) +
  geom_point() +
  labs(size = "year")
```



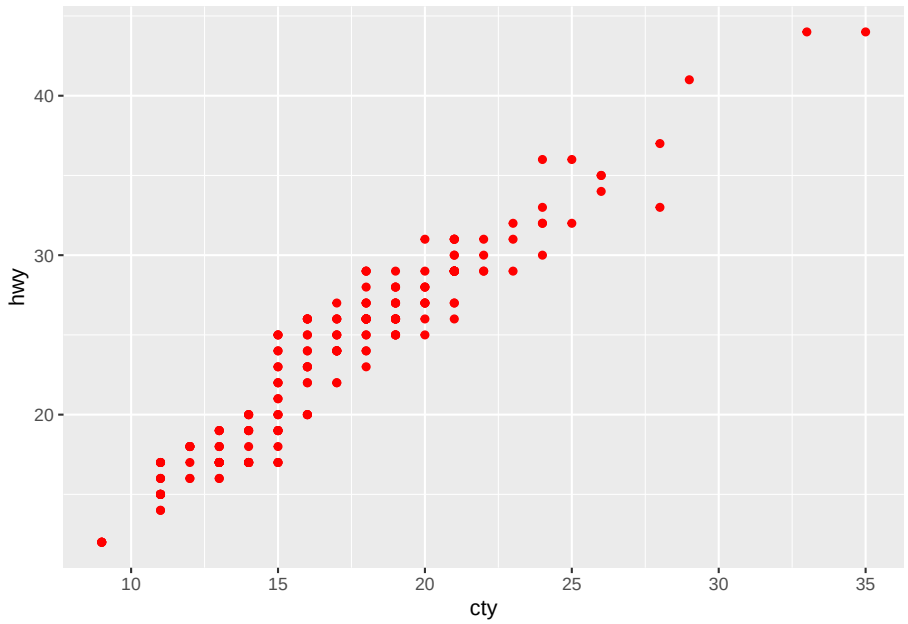

```
# Alternativa
ggplot(dados, aes(x = cty, y = hwy, col = cty <= 20)) +
  geom_point() +
  geom_vline(xintercept = 20)+
  labs(col = "year")
```



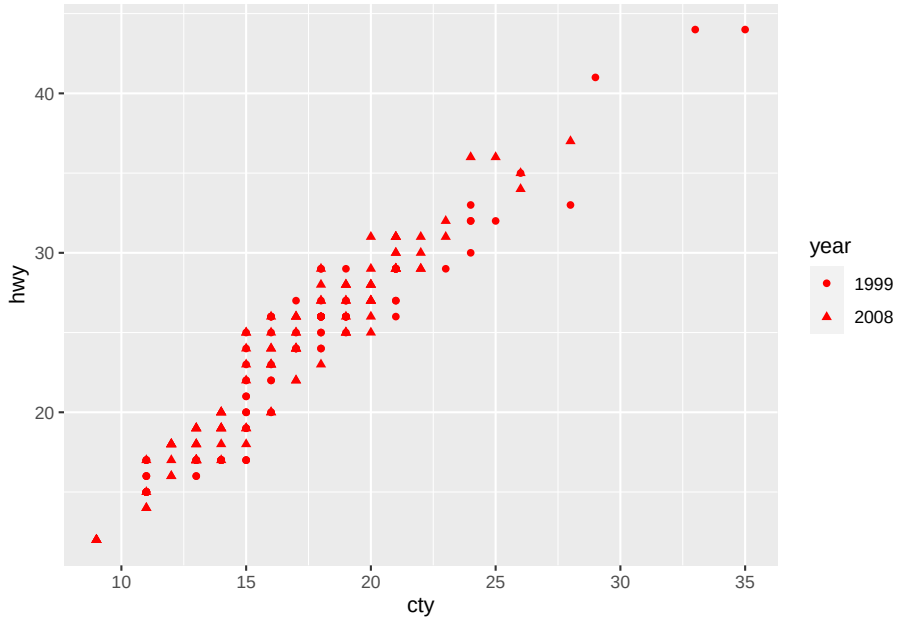

```
# Erro comum
ggplot(dados, aes(x = cty, y = hwy, col = "red")) +
  geom_point()+
  labs(col = "year")
```



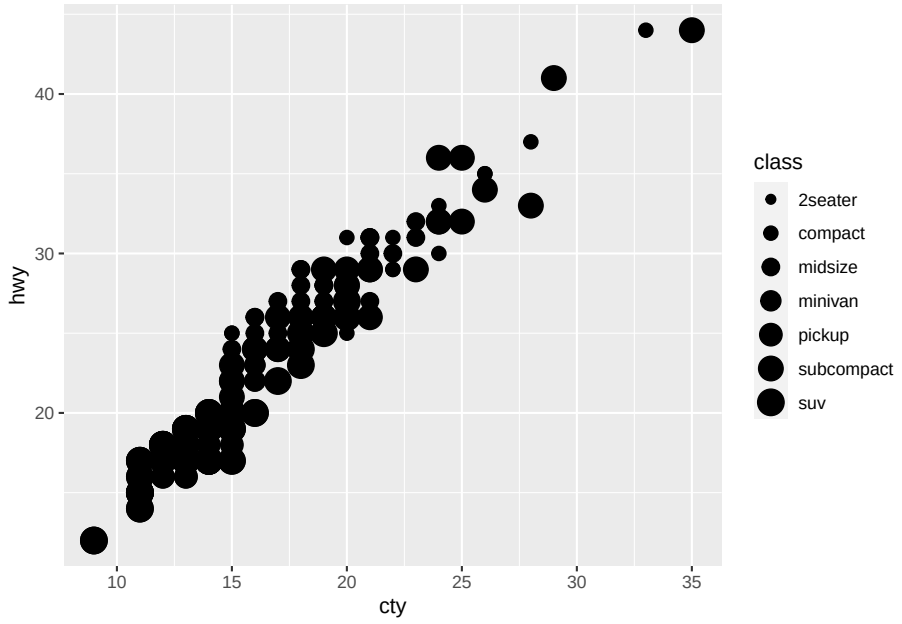
```
ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point(col = "red")+
  labs(col = "year")
```



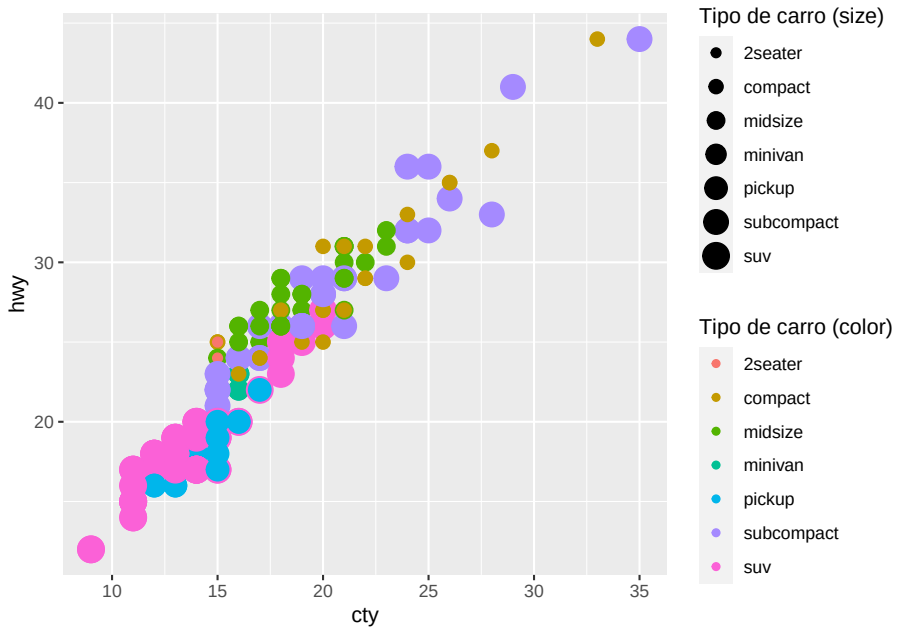

```
# Fim  Erro comum

ggplot(dados, aes(x = cty, y = hwy, shape = factor(year))) +
  geom_point(col = "red") +
  labs(shape = "year")
```



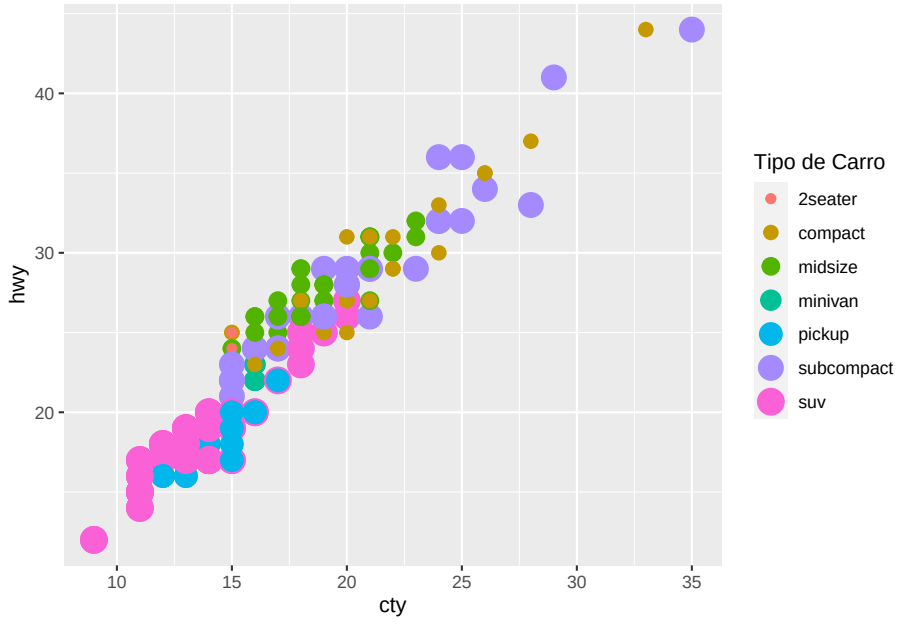

```
ggplot(dados, aes(x = cty, y = hwy, size = class)) +
  geom_point() +
  labs(size = "class")
```



```
ggplot(dados, aes(x = cty, y = hwy,
                  size = class,
                  col = class)) +
  geom_point() +
  guides(colour = guide_legend("Tipo de carro (color)"),
         size = guide_legend("Tipo de carro (size)"))
```

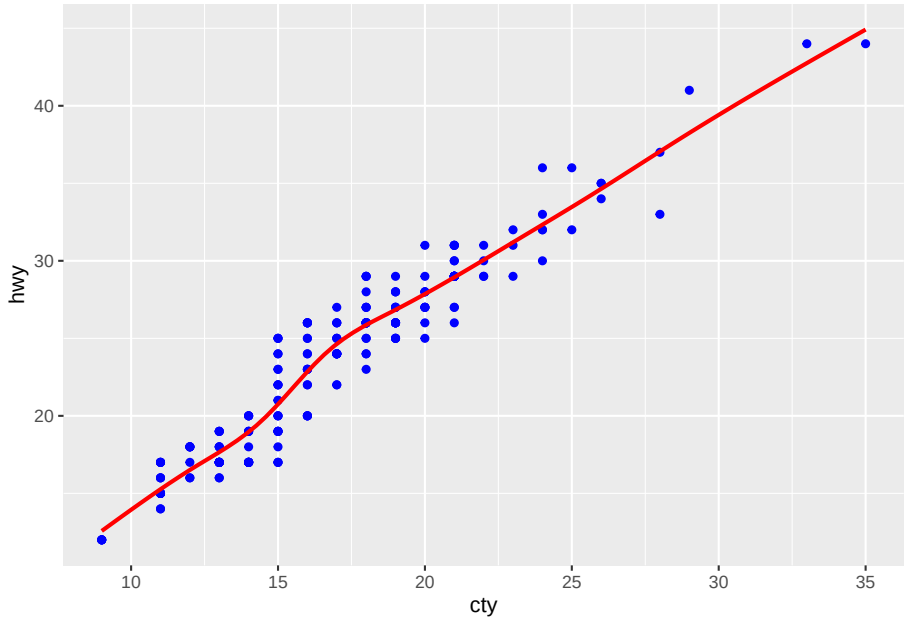



```
ggplot(dados, aes(x = cty, y = hwy,
                  size = class,
                  col = class)) +
  geom_point() +
  labs(col = "Tipo de Carro", size = "Tipo de Carro")+
  guides(col = "legend")
```

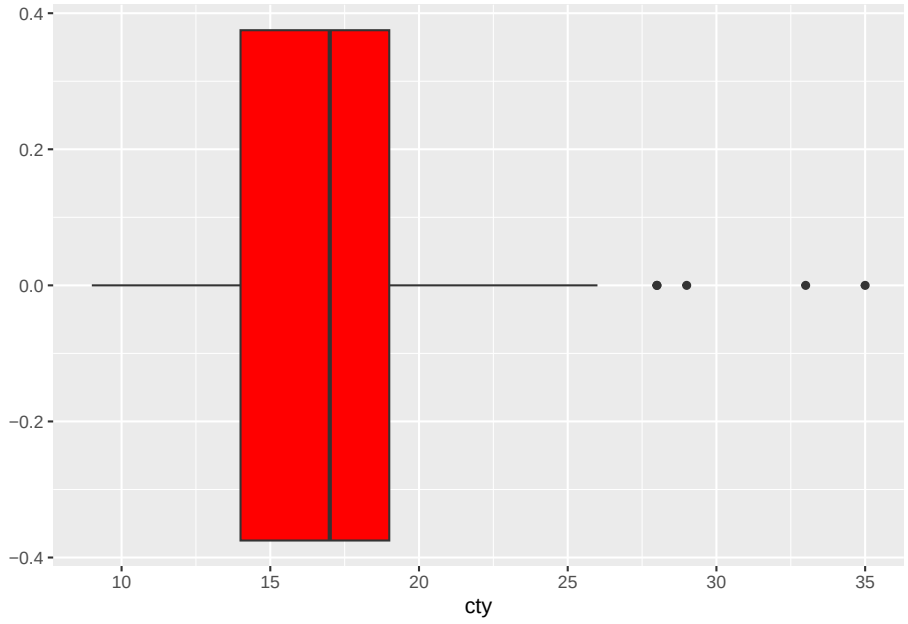


3.3 smooth, boxplot, histogram

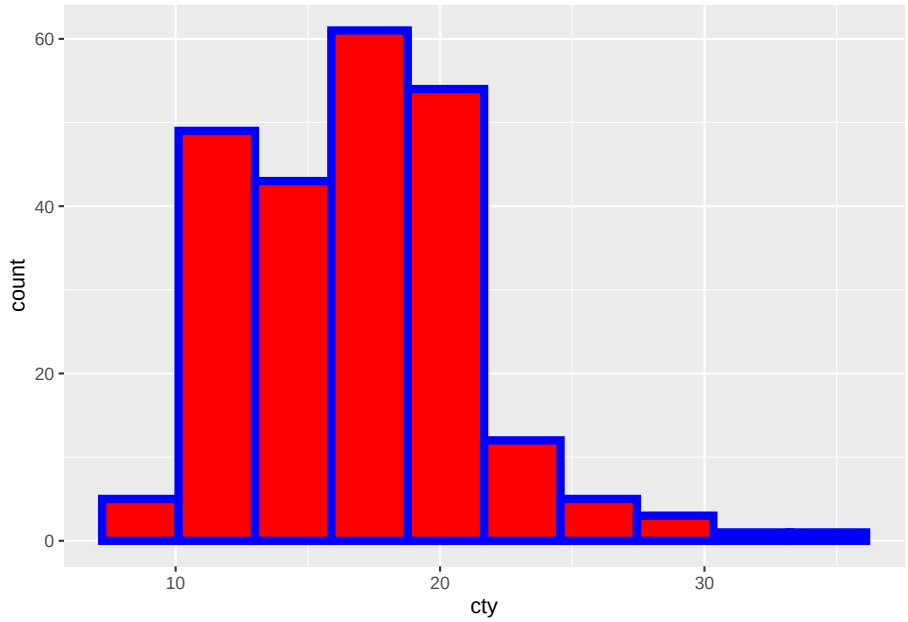
```
v1<- ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point(col = "blue")+
  geom_smooth(method = mgcv::gam,
              formula = y ~ s(x, bs = "cs") ,
              col = "red",
              se = FALSE)
v1
```



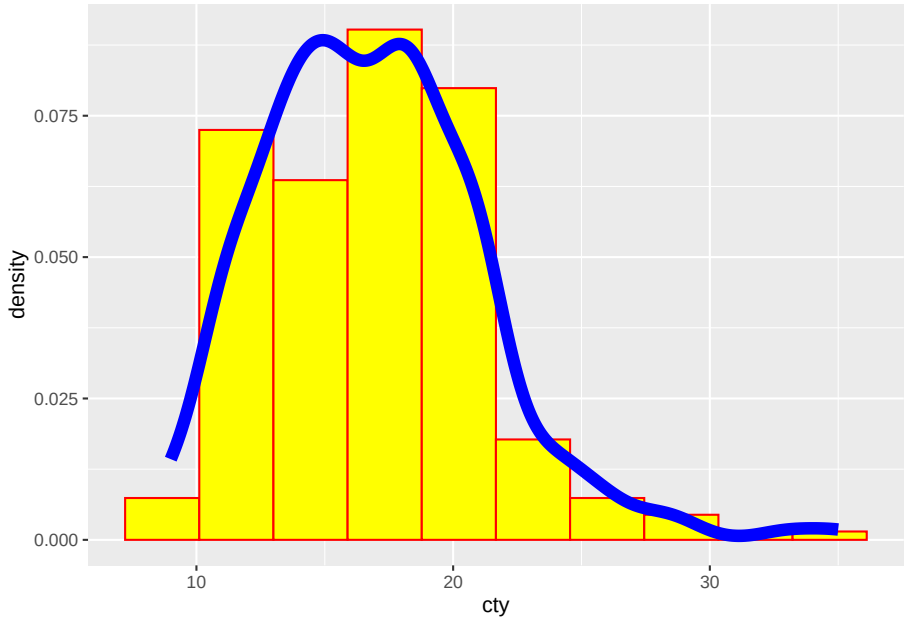

```
v2 <- ggplot(dados, aes(x = cty)) +
  geom_boxplot(fill = "red")
v2
```



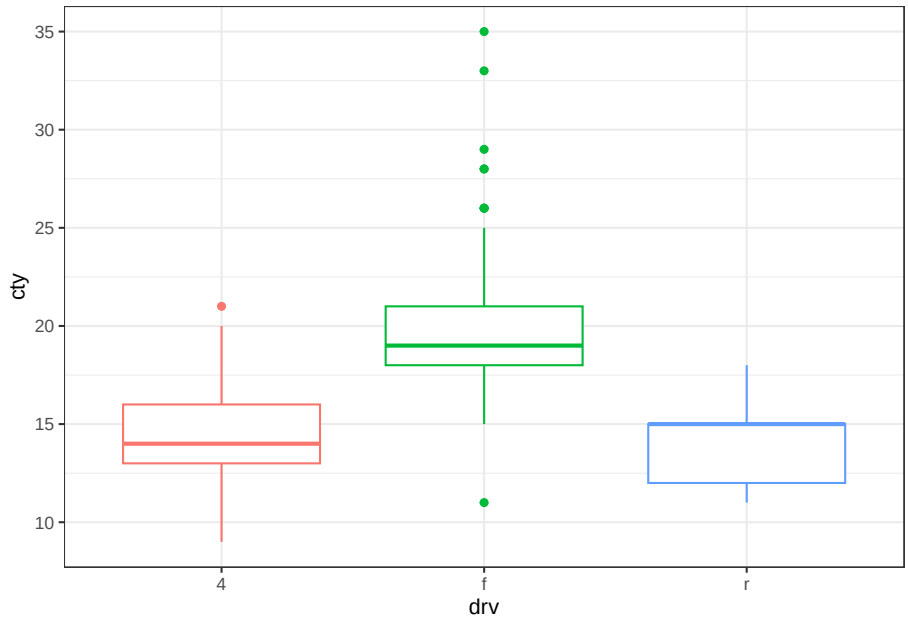
```
v3 <- ggplot(dados, aes(x = cty)) +
  geom_histogram(bins = 10, fill = "red", col = "blue", lwd=2)
v3
```




```
v4<- ggplot(dados, aes(x = cty)) +
  geom_histogram(aes(y = after_stat(density)),
                 bins = 10, fill = "yellow", col = "red") +
  geom_density(col = "blue", lwd =3)
v4
```




```
# Adicional (estatístic experimental)
ggplot(dados, aes(x = drv, y = cty, col = drv)) +
  geom_boxplot()+
  theme_bw()+
  theme(legend.position = "none")
```



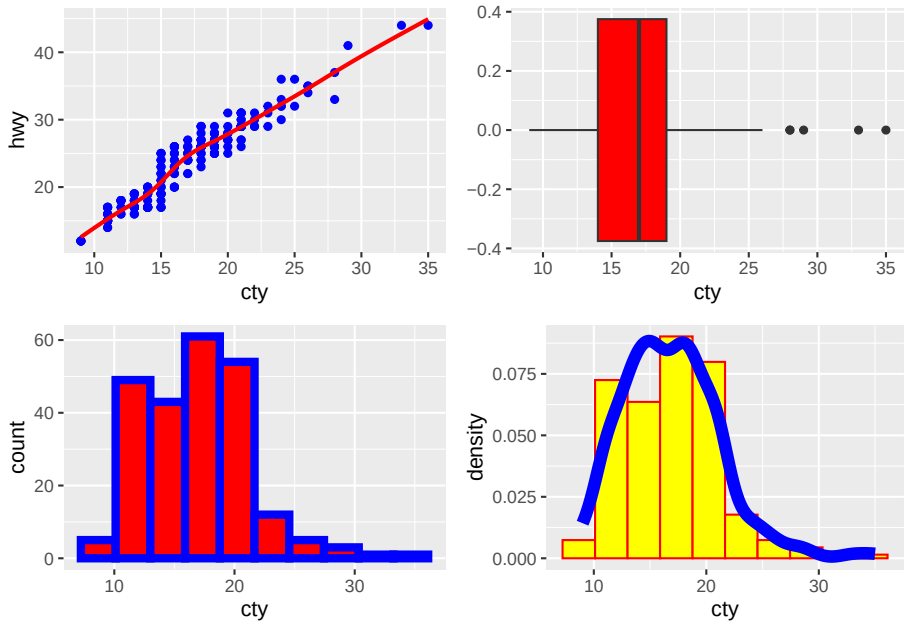
3.4 gridExtra e patchwork

Alguns links

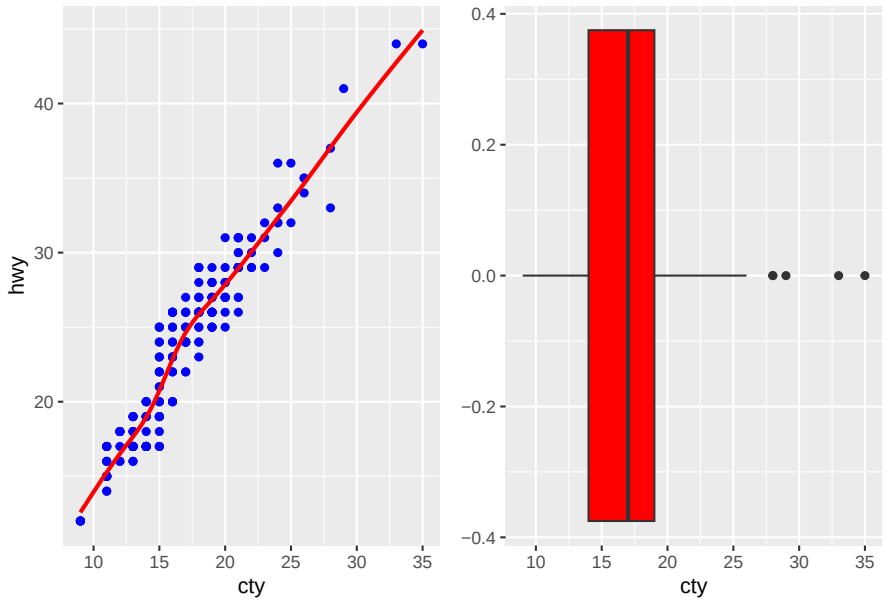
link 1: patchwork

link 2: patchwork

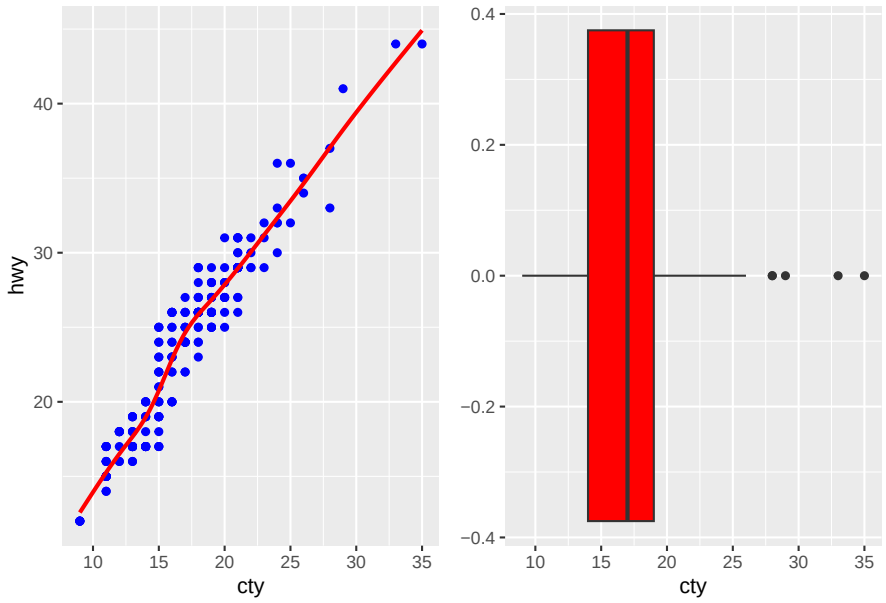
```
# gridExtra
grid.arrange(v1, v2, v3, v4)
```



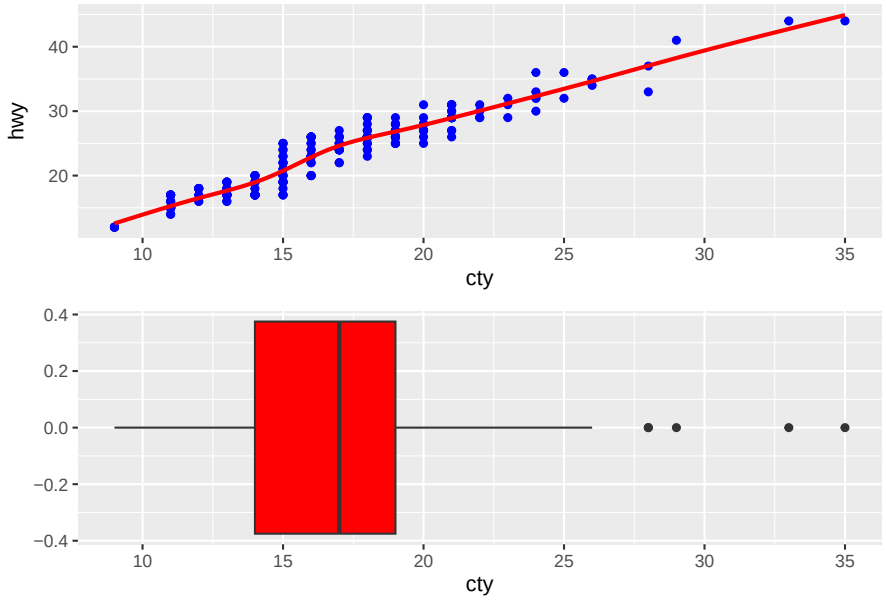

```
# patchwork
v1 + v2
```



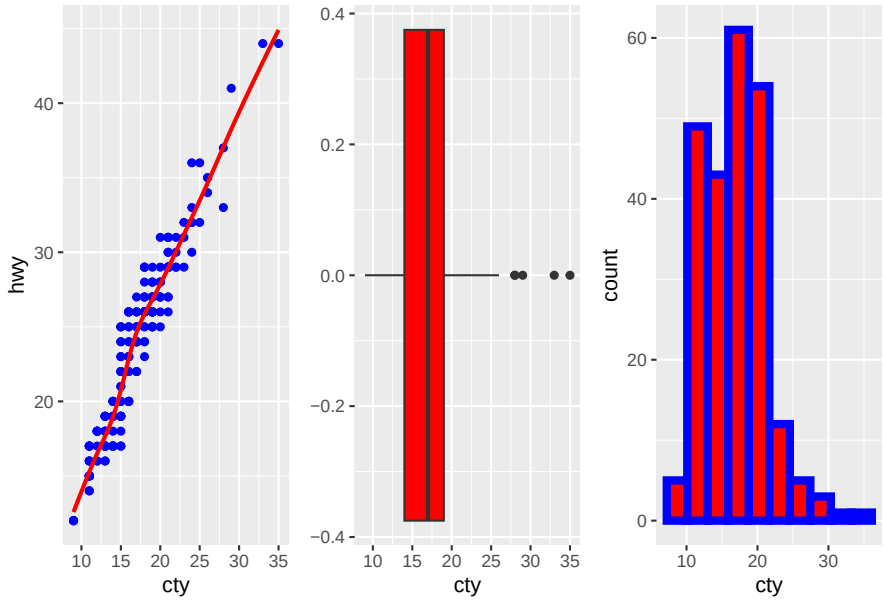
```
v1 | v2
```


hwy
cty
40
30
20
10
15
20
25
30
35
0.4
0.2
0.0
−0.2
−0.4

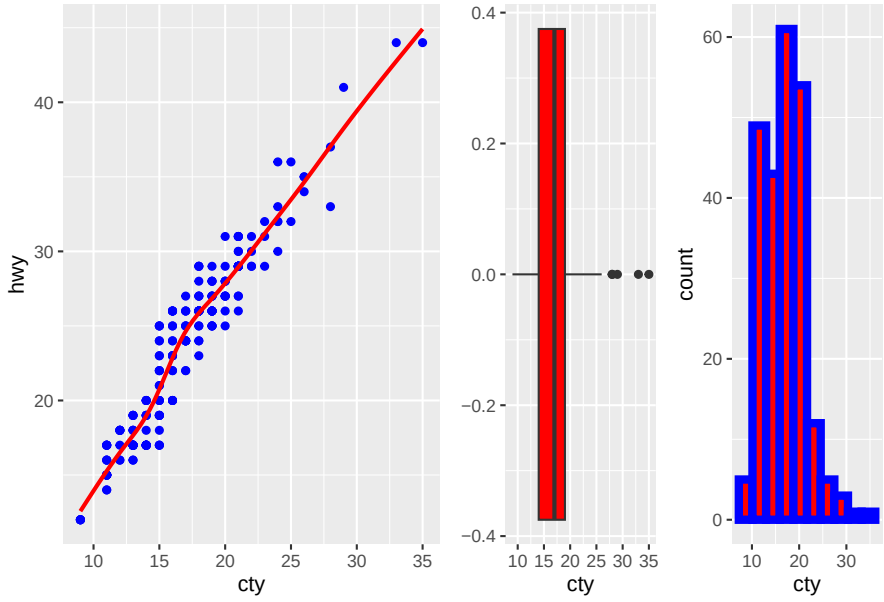
```
v1 / v2
```


hwy
cty
40
30
20
10
15
20
25
30
35
0.4
0.2
0.0
−0.2
−0.4

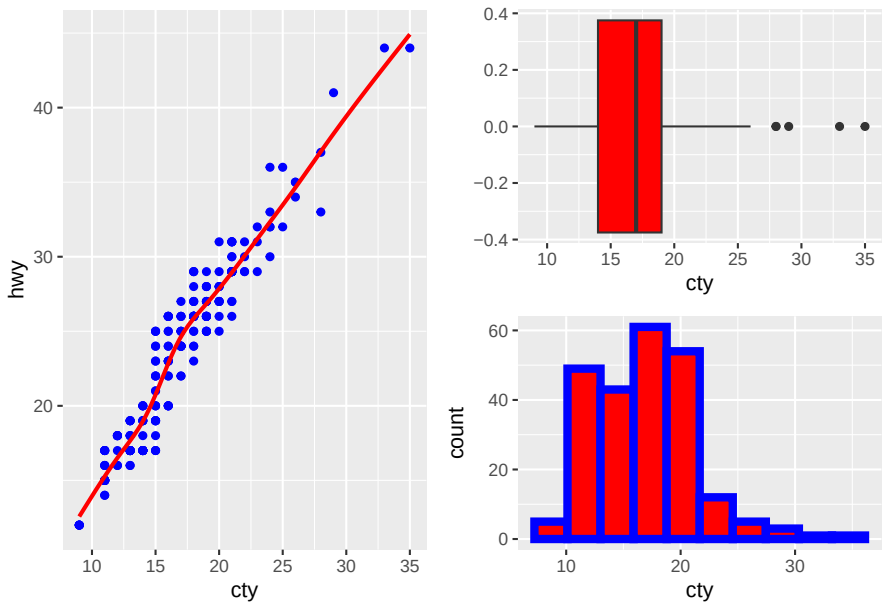
```
v1 + v2 + v3
```



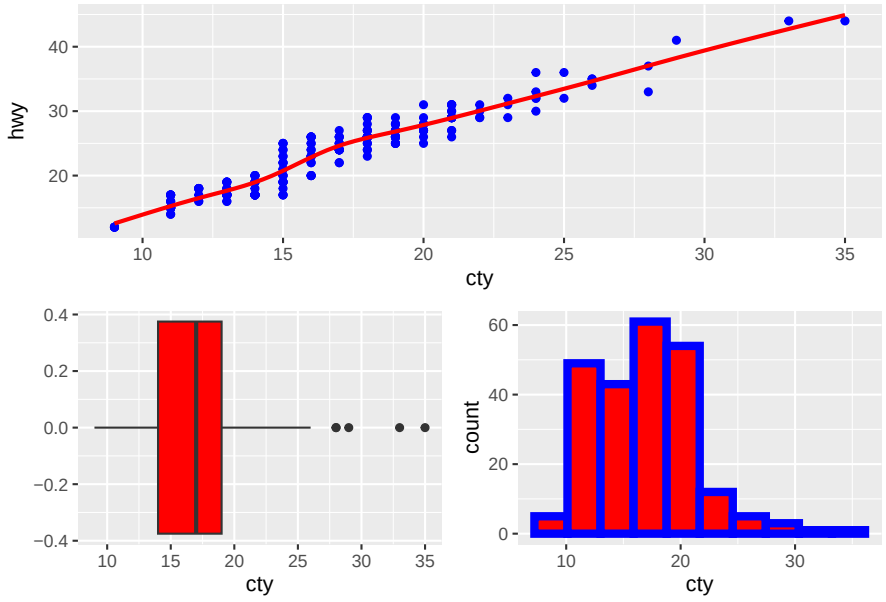

```
v1 + (v2 + v3)
```



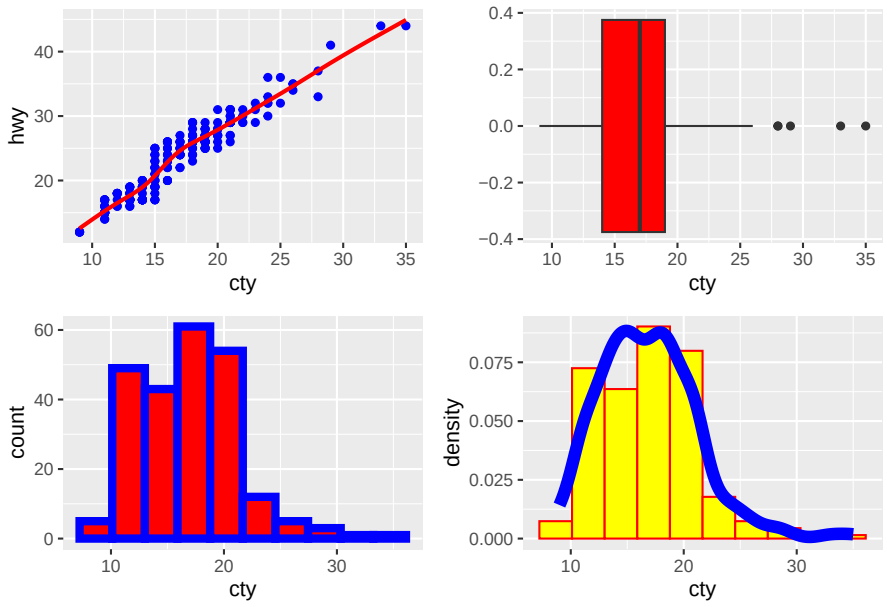
```
v1 | (v2 / v3)
```



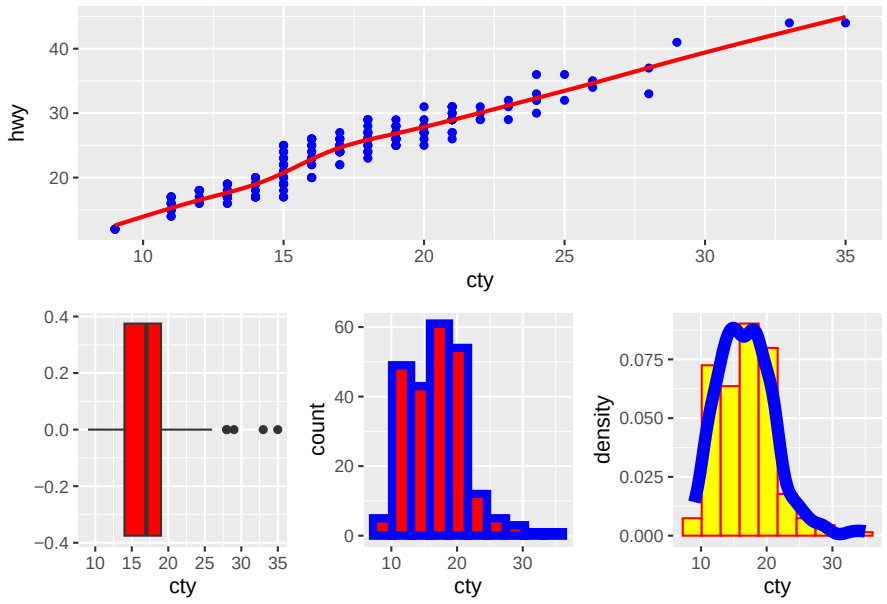

```
v1 / (v2 + v3)
```



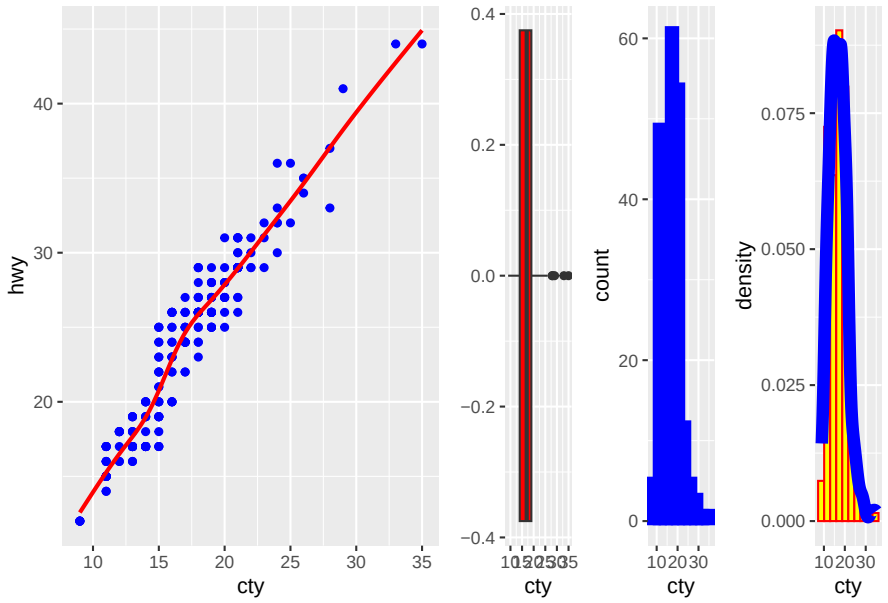
```
v1 + v2 + v3 + v4
```



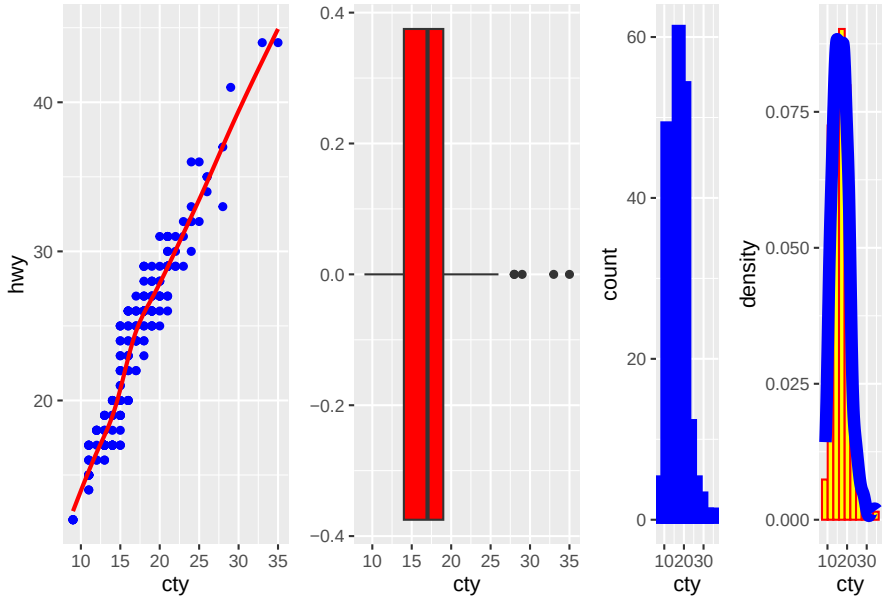

```
v1/(v2+v3+v4)
```



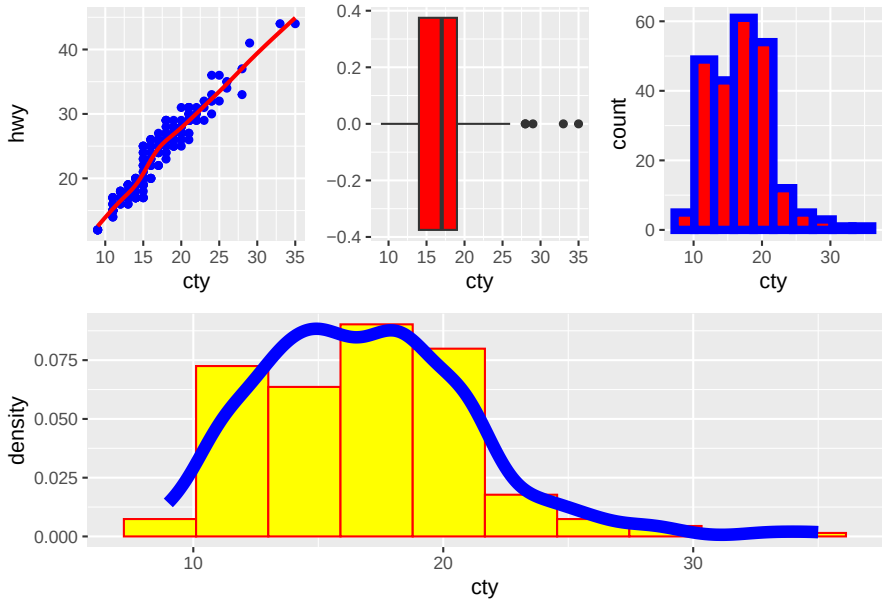
```
v1  + (v2 + v3 + v4)
```




```
v1  + v2 + (v3 + v4)
```

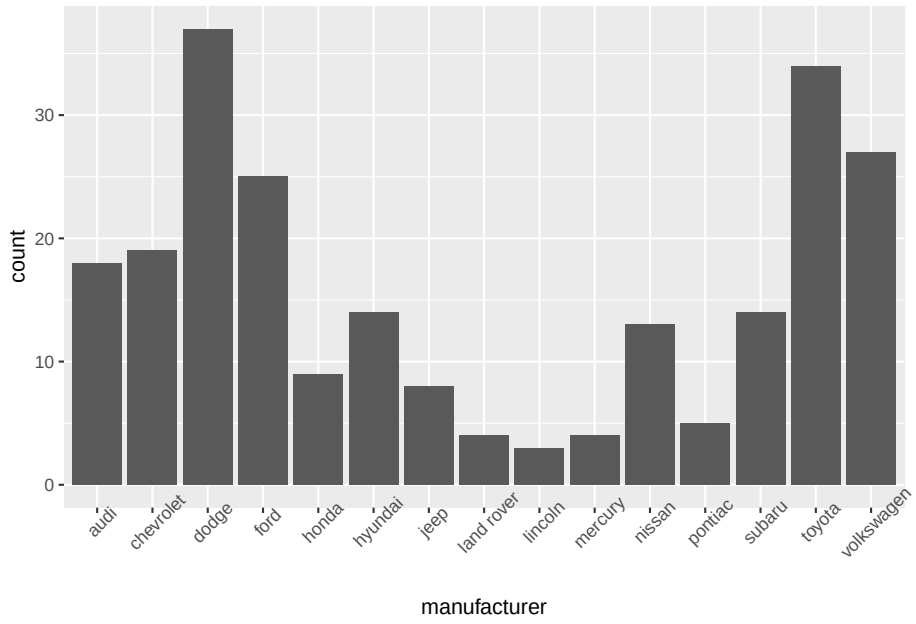


```
(v1 | v2 | v3) / v4
```

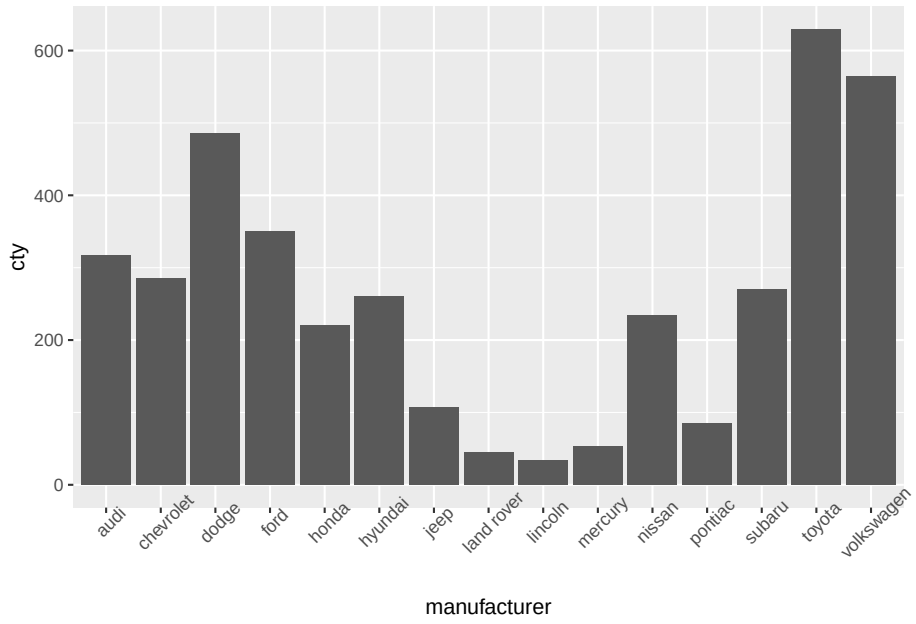


3.5 bar, col, density, density2d

```
v5 <- ggplot(dados , aes(x = manufacturer)) +
  geom_bar()+
  theme(axis.text.x = element_text(angle = 45))
v5
```



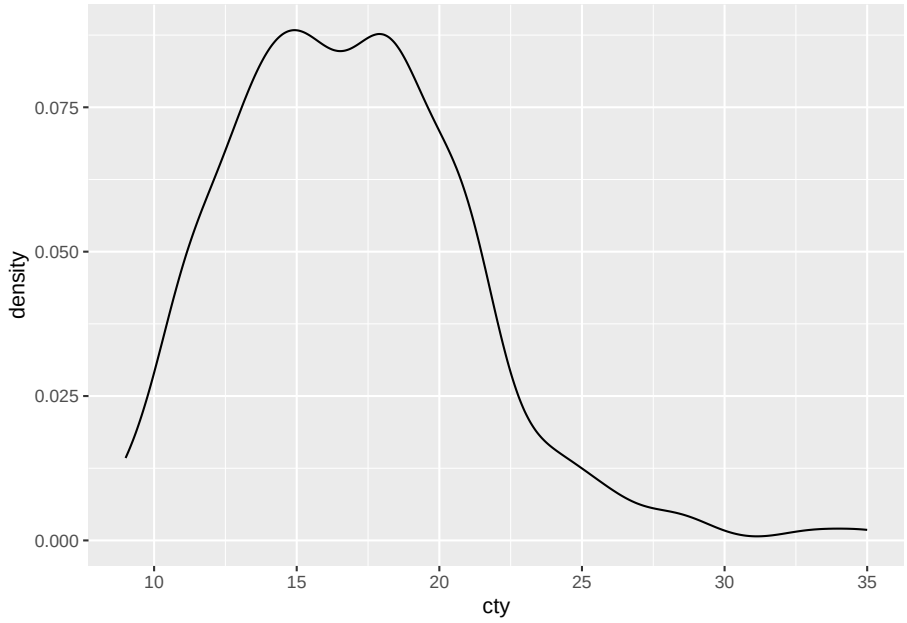

```
# Dúvidas no geom_col
v6 <- ggplot(dados , aes(x = manufacturer, y = cty)) +
  geom_col()+
  theme(axis.text.x = element_text(angle = 45))
v6
```



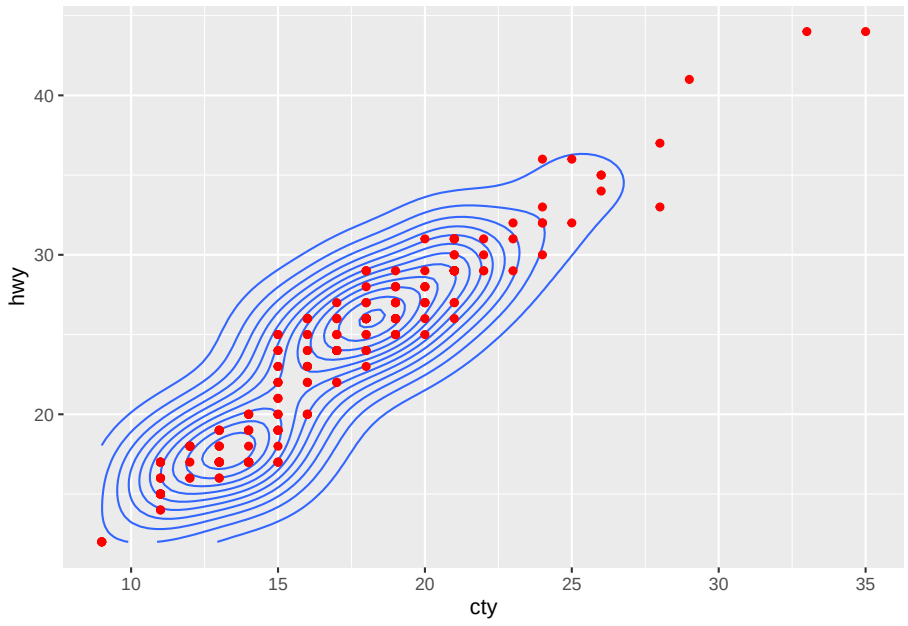

```
dados %>%
  select(manufacturer, cty) %>%
  group_by(manufacturer) %>%
  summarise(soma_total_cty = sum(cty),
            n = n())
```

```
## # A tibble: 15 x 3
##    manufacturer soma_total_cty     n
##    <fct>                 <int> <int>
##  1 audi                    317    18
##  2 chevrolet               285    19
##  3 dodge                   486    37
##  4 ford                    350    25
##  5 honda                   220     9
##  6 hyundai                 261    14
##  7 jeep                    108     8
##  8 land rover               46     4
##  9 lincoln                  34     3
## 10 mercury                  53     4
## 11 nissan                  235    13
## 12 pontiac                  85     5
## 13 subaru                  270    14
## 14 toyota                  630    34
## 15 volkswagen              565    27
```

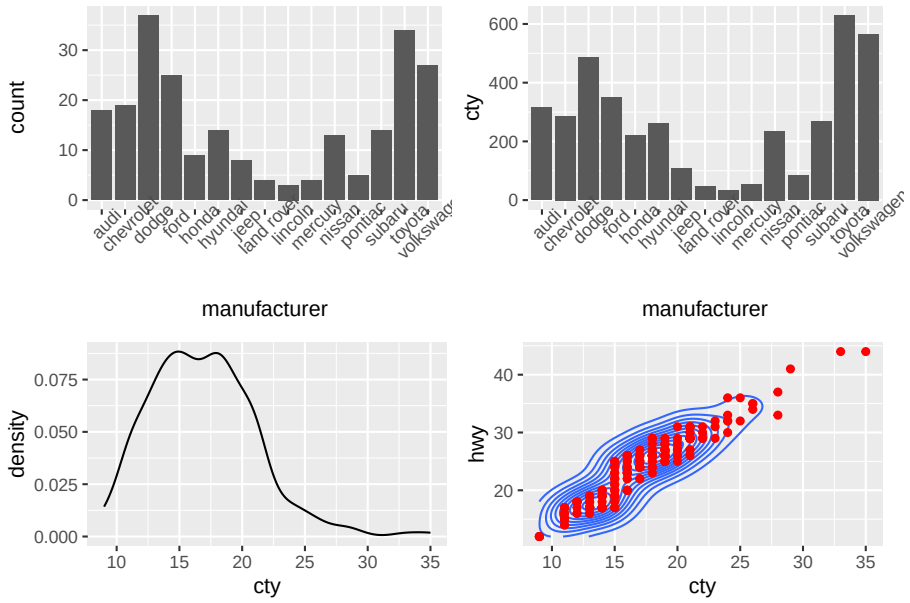
```
# dados %>%
#   filter(manufacturer == "audi") %>%
#   select(cty) %>%
#   sum()
v7 <- ggplot(dados , aes(x = cty)) +
  geom_density()
v7
```




```
v8 <- ggplot(dados, aes(x = cty, y = hwy)) +
  geom_density2d()+
  geom_point(colour = "red")
v8
```


hwy
cty
40
30
20
10
15
20
25
30
35

```
(v5+v6)/ (v7 + v8)
```


count
cty
manufacturer
manufacturer
density
hwy
cty
cty
audi
chevrolet
dodge
ford
honda
hyundai
jeep
land rover
lincoln
mercury
nissan
pontiac
subaru
toyota
volkswagen

```
# Deixar pra depois...
 dados %>%
    select(manufacturer, hwy, year) %>%
    filter(manufacturer == "audi", year == "1999") %>%
    summarise(media = max(hwy))
```

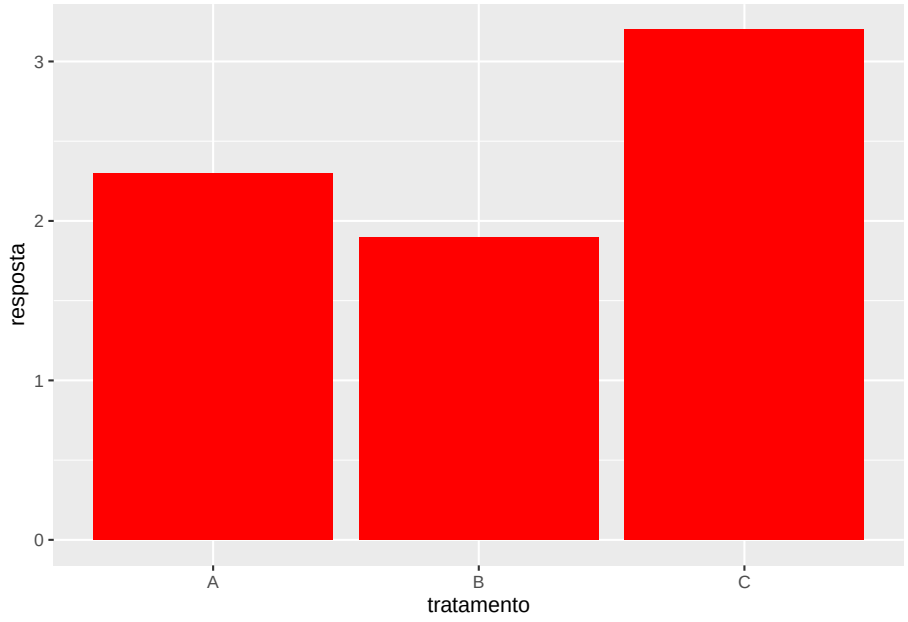
```
## # A tibble: 1 x 1
##   media
##   <int>
## 1    29
```

```
# plotly
ggplotly(
ggplot(dados, aes(x = manufacturer, y = hwy, fill = factor(year))) +
  geom_col(position = "dodge") +
  labs(fill = "year") +
  theme(axis.text.x = element_text(angle = 45)))

dados %>% select(manufacturer, hwy, year) %>%
  group_by(manufacturer, year) %>%
  summarise(media = mean(hwy))
```

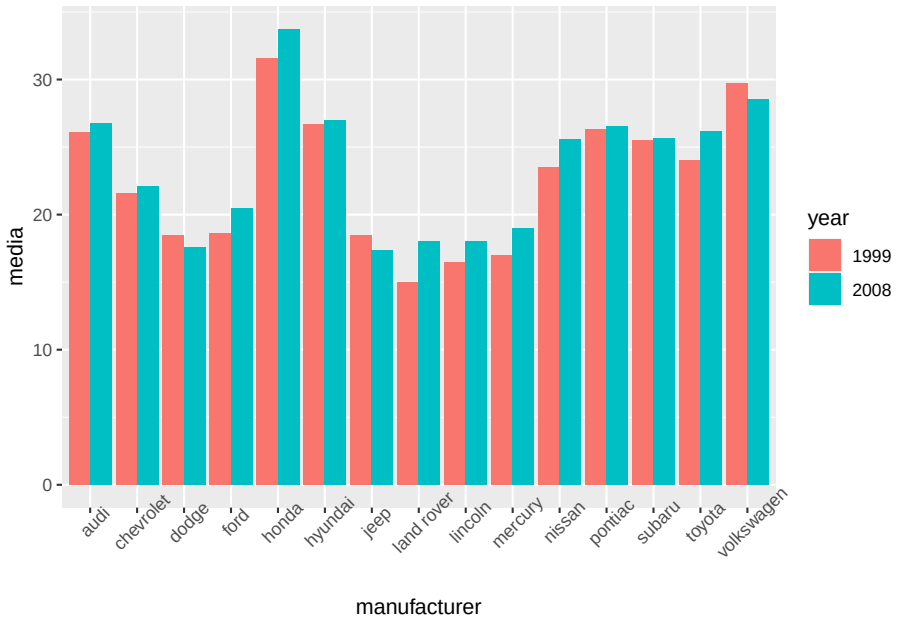
```
# Para pensar

(dados_trat <- data.frame(tratamento = LETTERS[1:3],
                          resposta = c(2.3, 1.9, 3.2)))
```

```
##   tratamento resposta
## 1          A      2.3
## 2          B      1.9
## 3          C      3.2
```

```
ggplot(dados_trat, aes(tratamento, resposta)) +
  geom_col(fill = "red")
```

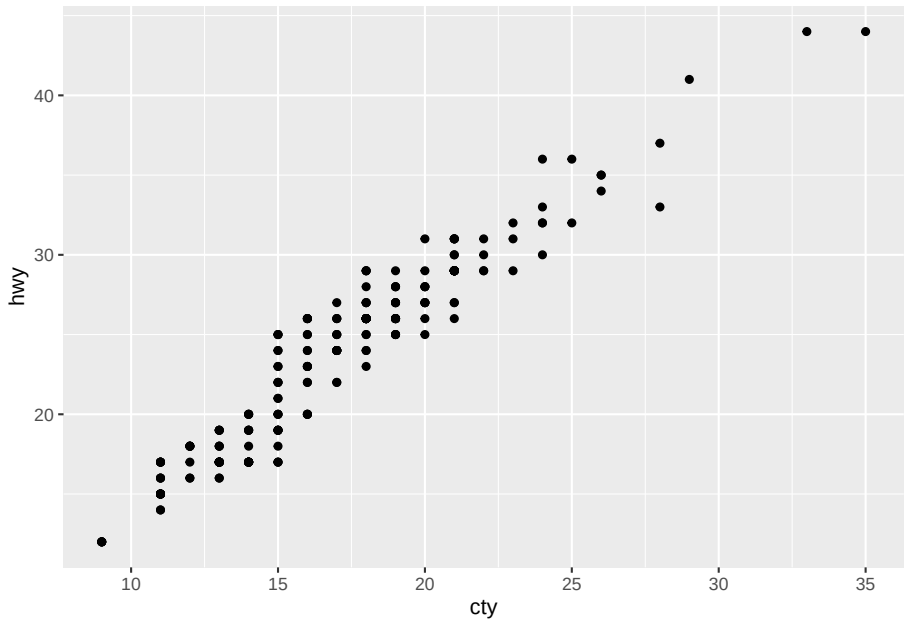



```
# Mais detalhes...
dados %>% select(manufacturer, hwy, year) %>%
  group_by(manufacturer, year) %>%
  summarise(media = mean(hwy), .groups = "drop") %>%
  ggplot(aes(x = manufacturer, y = media, fill = factor(year)))+
  geom_col(position = "dodge")+
  labs(fill = "year") +
  theme(axis.text.x = element_text(angle = 45))
```

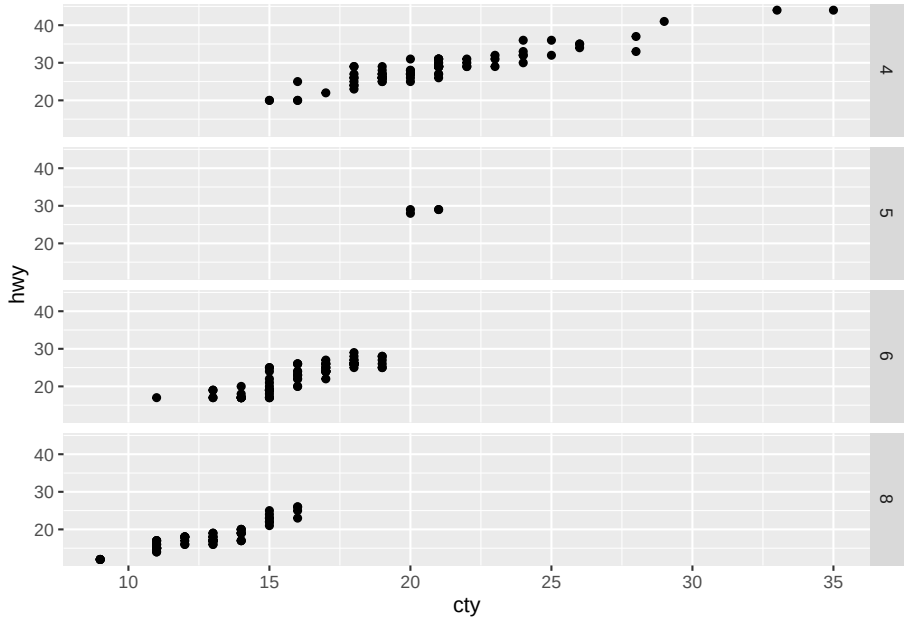


3.6 facet_grid, facet_wrap

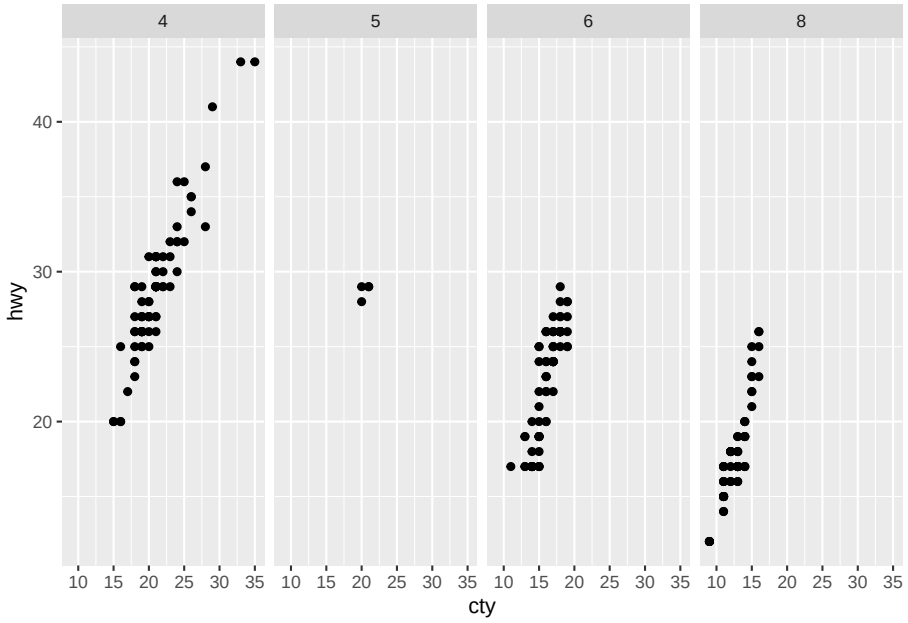
```
p1<- ggplot(dados, aes(x = cty, y = hwy)) +
  geom_point()
p1
```



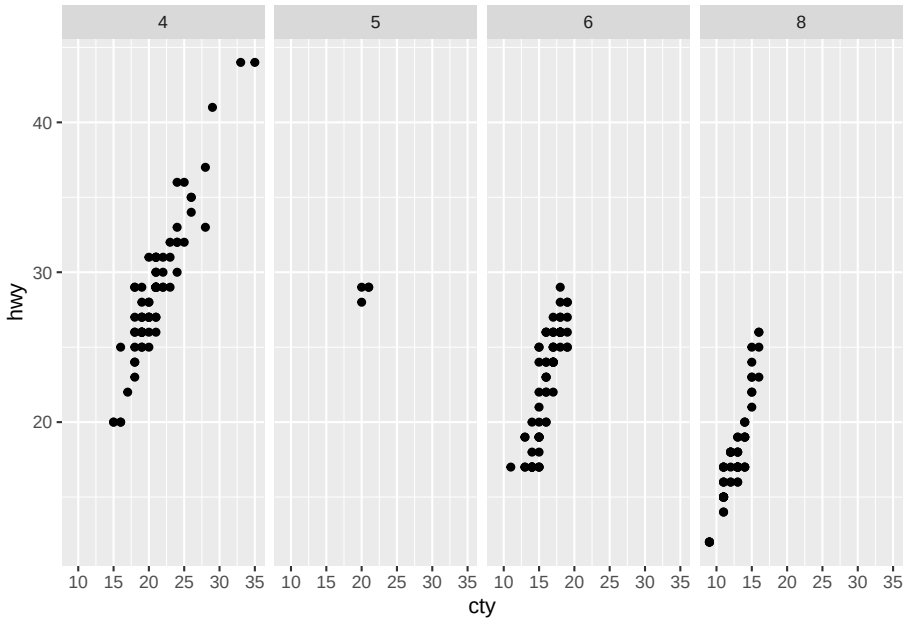

```
p1 + facet_grid(rows = vars(cyl))
```



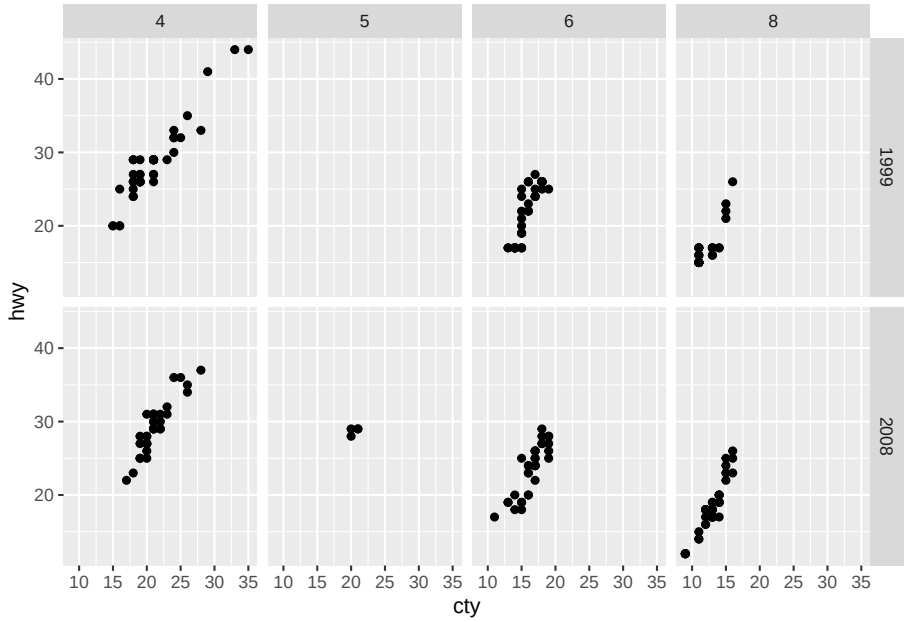
```
p1 + facet_grid(cols = vars(cyl))
```



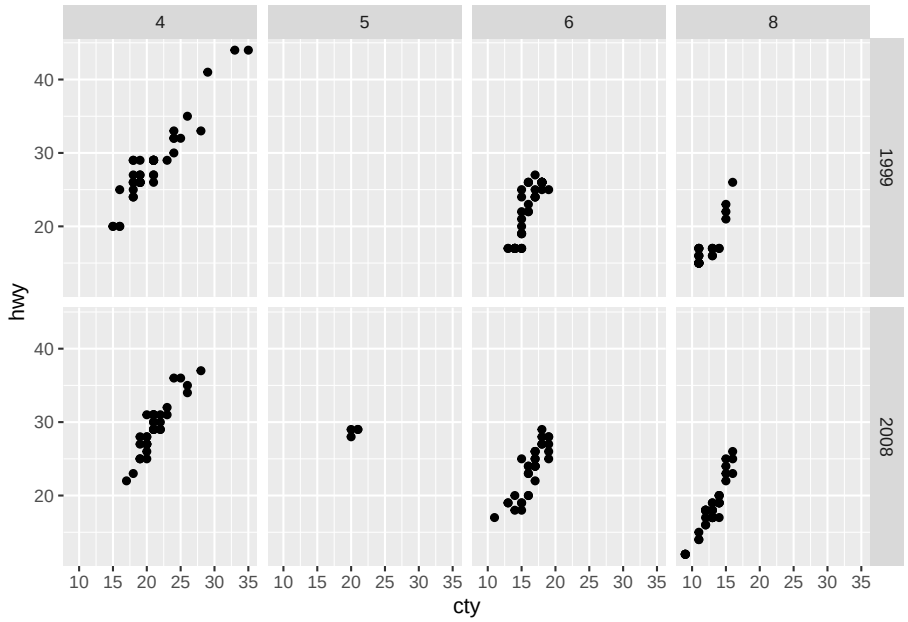

```
p1 + facet_grid(~cyl)
```



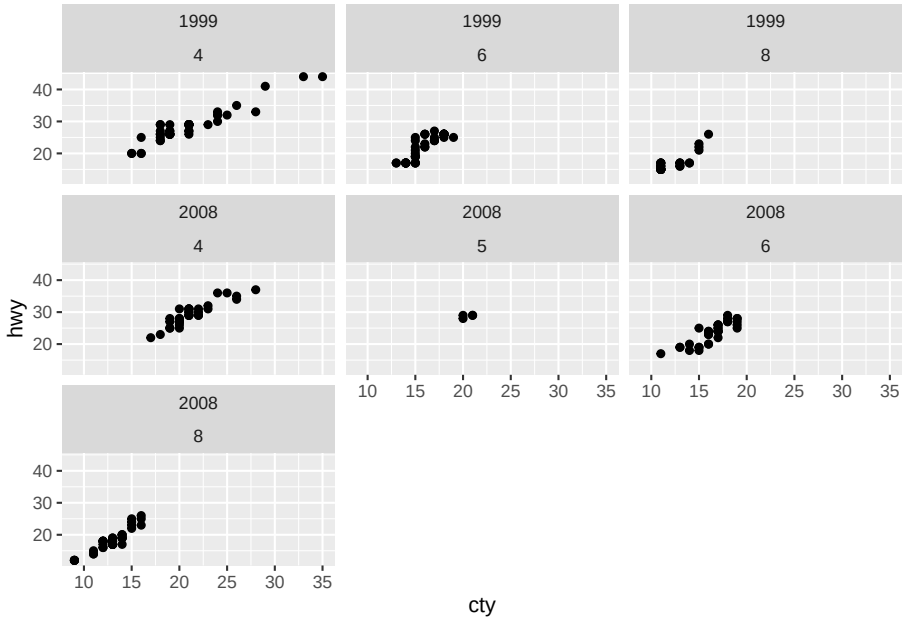
```
p1 + facet_grid(rows = vars(year), cols =vars(cyl))
```



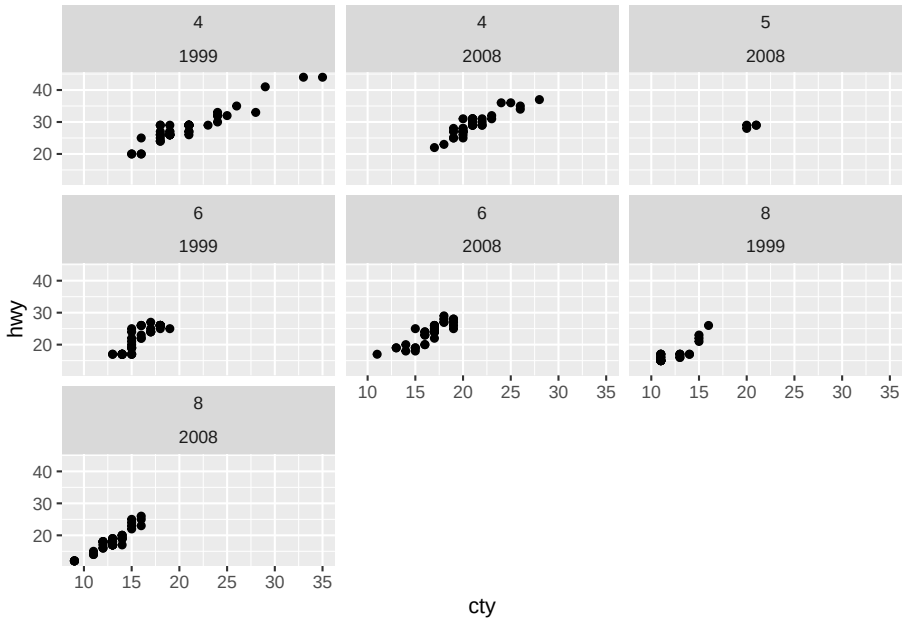

```
p1 + facet_grid(year~cyl)
```



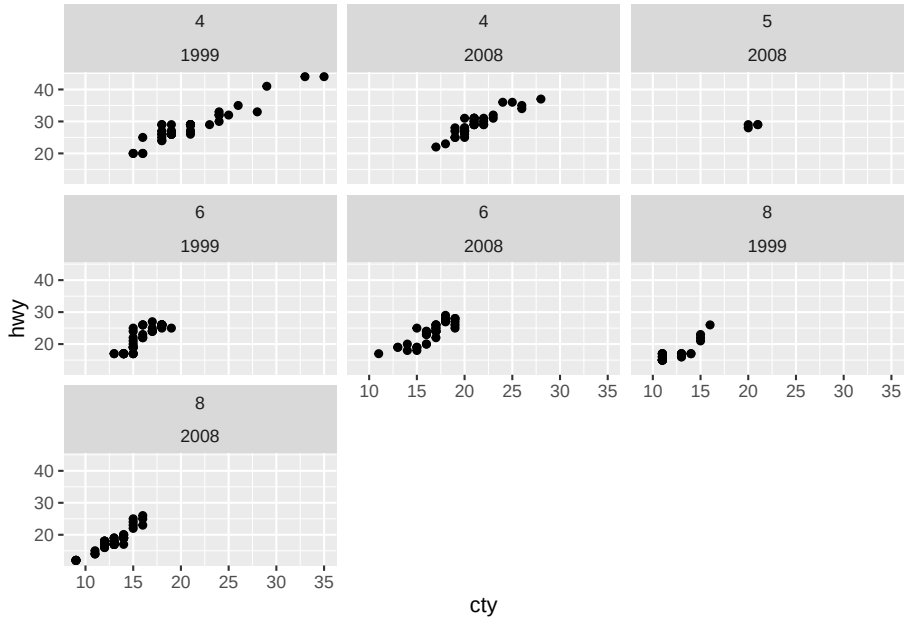
```
p1 + facet_wrap(year ~ cyl)
```



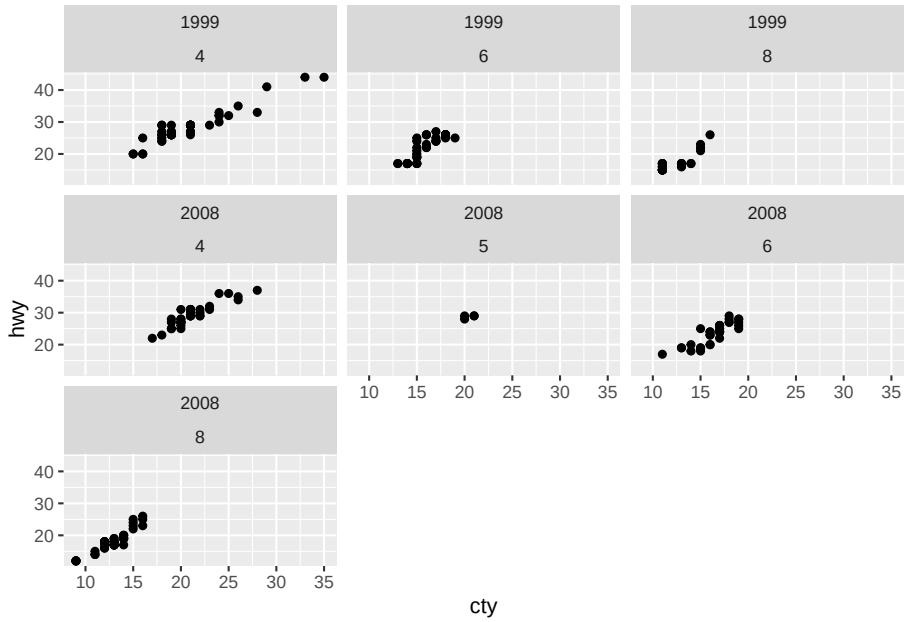

```
p1 + facet_wrap(cyl ~ year)
```



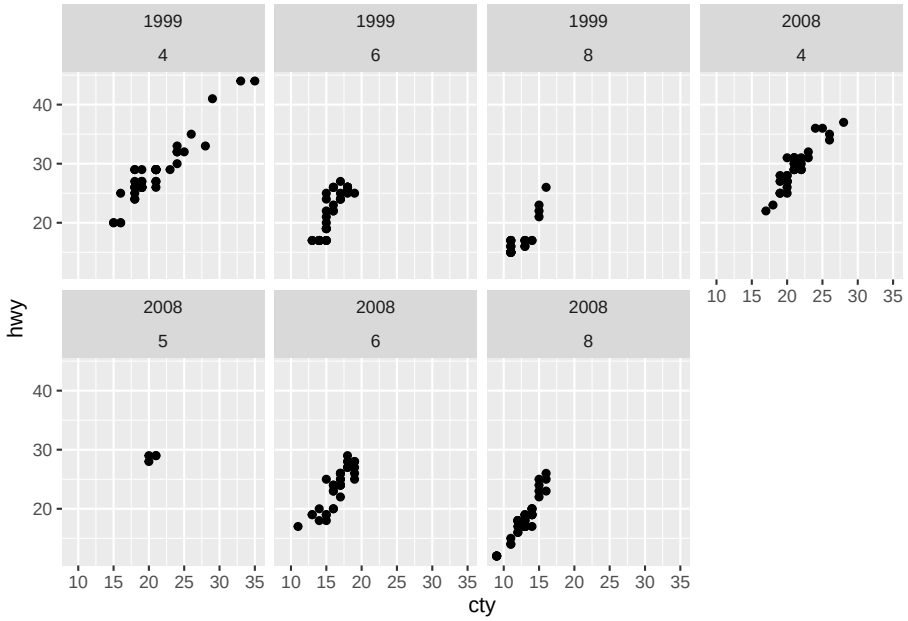
```
p1 + facet_wrap(~cyl + year)
```



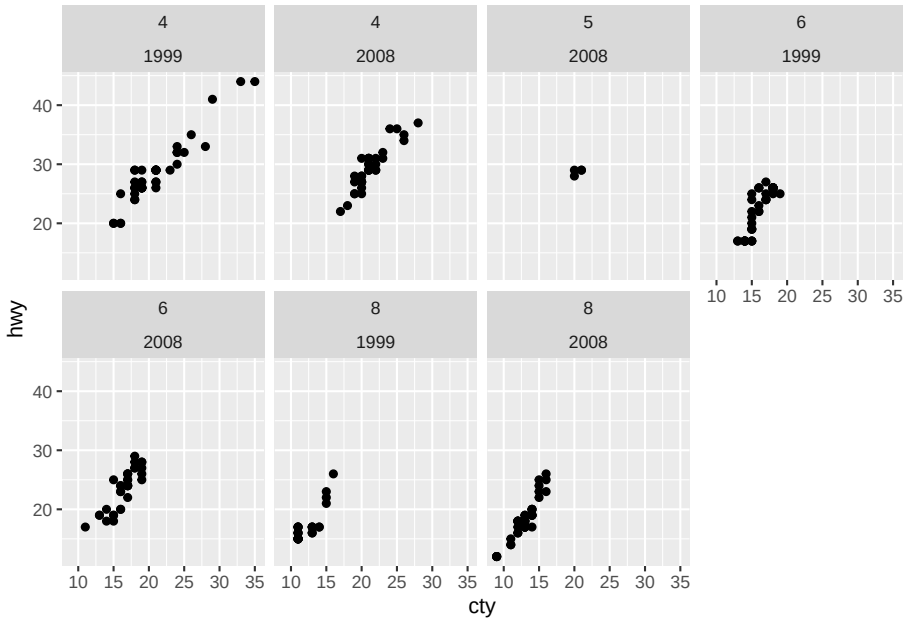

```
p1 + facet_wrap(~year + cyl)
```



```
p1 + facet_wrap(year ~ cyl, ncol = 4)
```

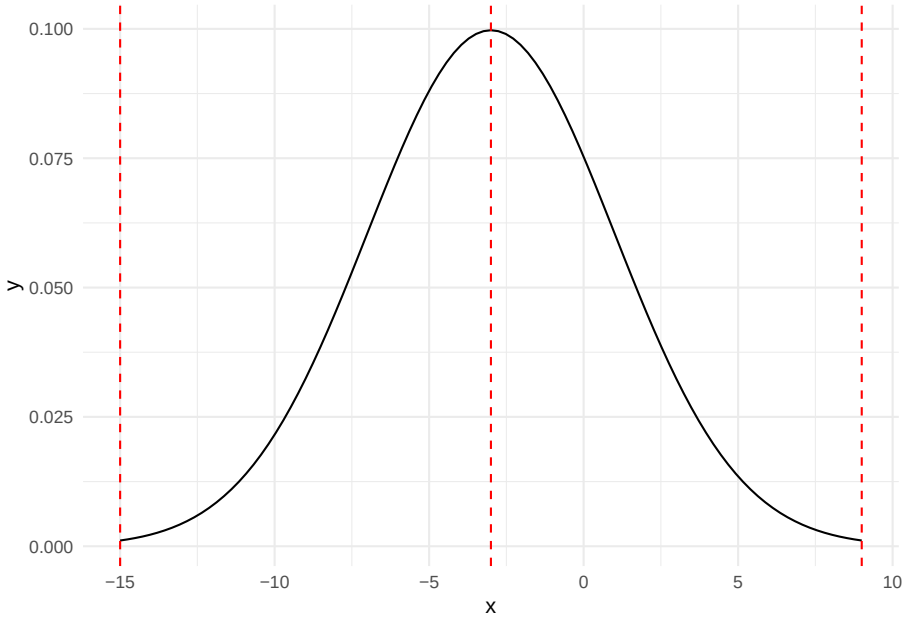



```
p1 + facet_wrap(cyl ~ year, ncol = 4)
```



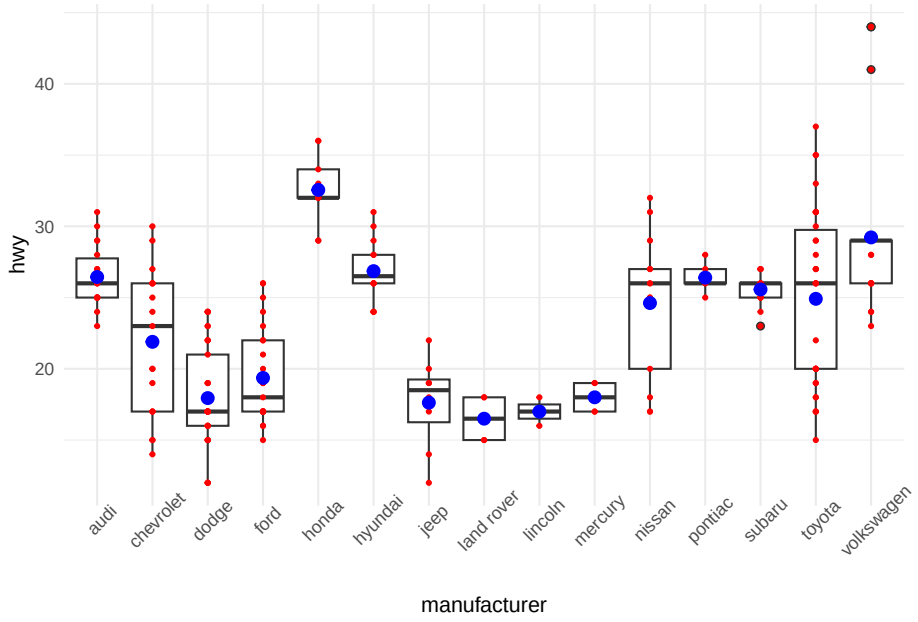
3.7 stat_function

```
a<- -3 # média
b<- 4  # desv. padrão
ggplot(data.frame(x = c(a - 3*b, a + 3*b)), aes(x)) +
  stat_function(fun = dnorm, args = list(mean = a, sd = b))+
  geom_vline(xintercept = c(a - 3*b, a, a + 3*b), col = "red", lty = 2)+
  theme_minimal()
```



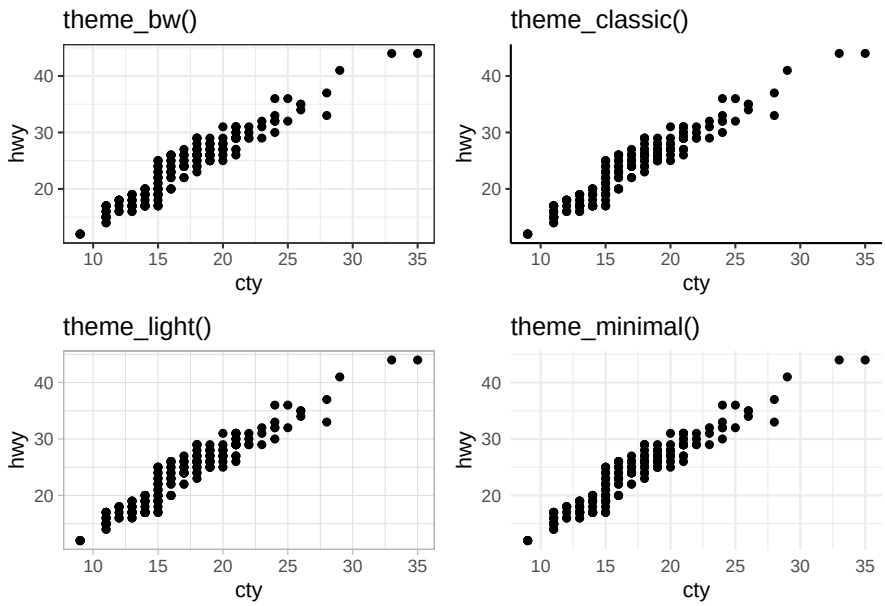
3.8 stat_summary

```
ggplot(dados, aes(x = manufacturer, y = hwy)) +
  geom_boxplot()+
  geom_point(col = "red", size=0.8)+
  stat_summary(fun = mean, col = "blue")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 45))
```



3.9 theme_*()

```
a1<- p1 + theme_bw() + labs(title = "theme_bw()")
a2<- p1 + theme_classic() + labs(title = "theme_classic()")
a3<- p1 + theme_light() + labs(title = "theme_light()")
a4<- p1 + theme_minimal() + labs(title = "theme_minimal()")

a1 + a2 + a3 + a4
```

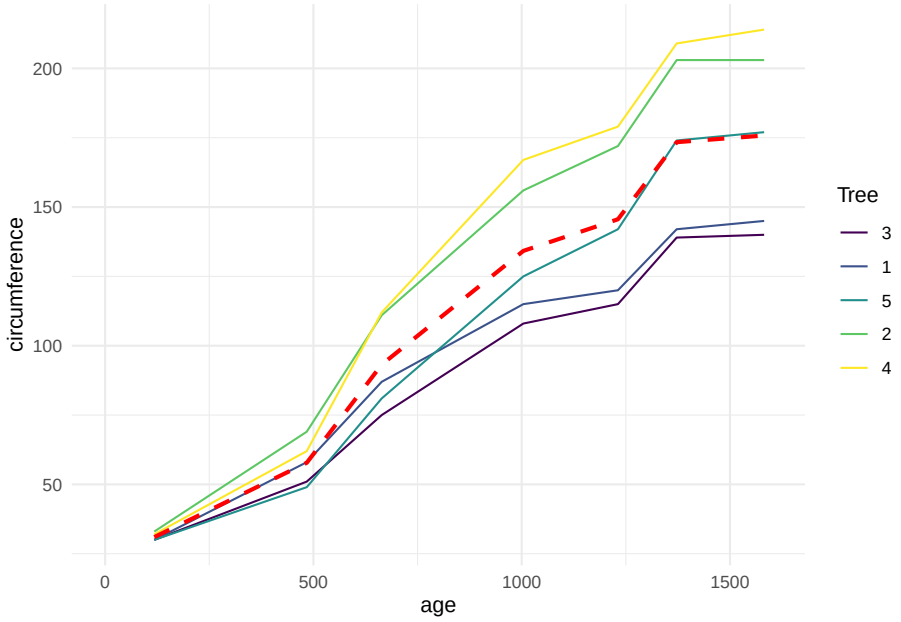


3.10 Gráfico de perfis (Spaguetti plot)

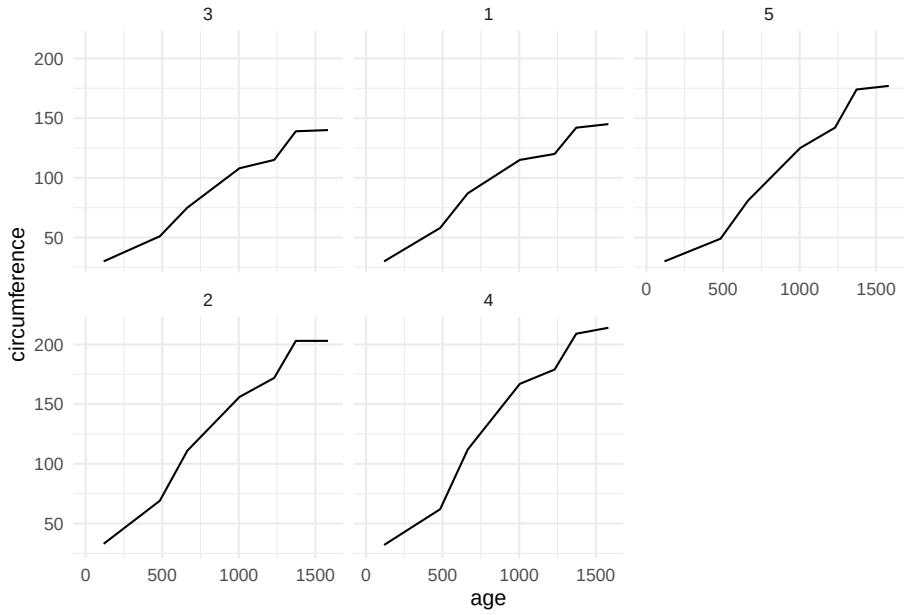
```
glimpse(Orange)
```

```
## Rows: 35
## Columns: 3
## $ Tree          <ord> 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3,~
## $ age           <dbl> 118, 484, 664, 1004, 1231, 1372, 1582, 118, 484, 664, 10~
## $ circumference <dbl> 30, 58, 87, 115, 120, 142, 145, 33, 69, 111, 156, 172, 2~
```

```
ggplot(Orange, aes(x = age, y = circumference, group = Tree,
                   col = Tree)) +
  geom_line()+
  stat_summary(aes(group = 1), fun = mean, col = "red",
               geom = "line", size = 1, show.legend = FALSE,
               linetype = 2)+
  xlim(0, 1600)+
  theme_minimal()
```




```
ggplot(Orange, aes(x = age, y = circumference, group = Tree)) +
  geom_line()+
  xlim(0, 1600)+
  facet_wrap(~Tree)+
  theme_minimal()+
  theme(legend.position = "none")
```



3.11 plotly

plotly cran

Interactive web-based data visualization with R, plotly, and shiny

Plotly R Open Source Graphing Library

```
ggplotly(v1)
ggplotly(v2)
ggplotly(v4)
ggplotly(v5)
```

3.12 esquisse

Alguns links de interesse

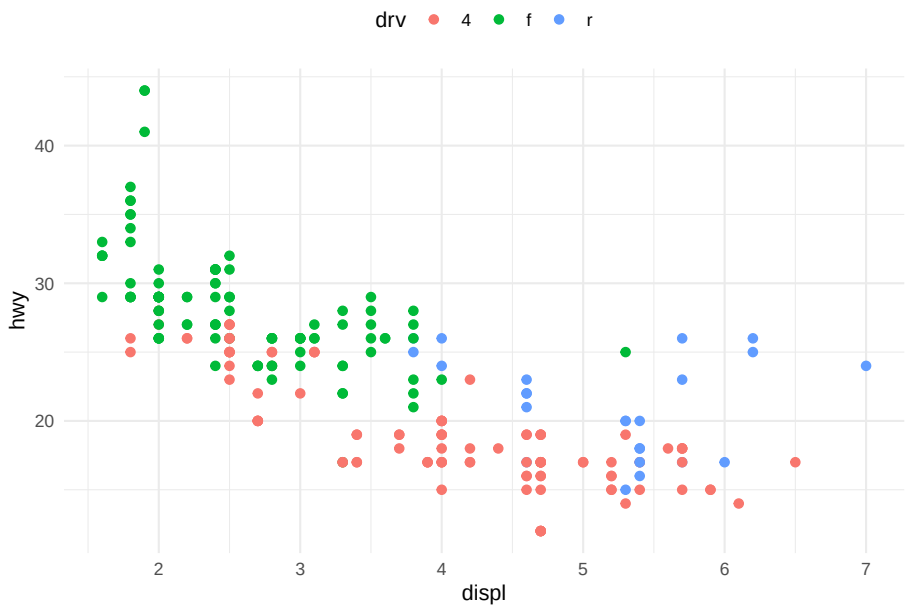
esquisse

esquisse + shiny

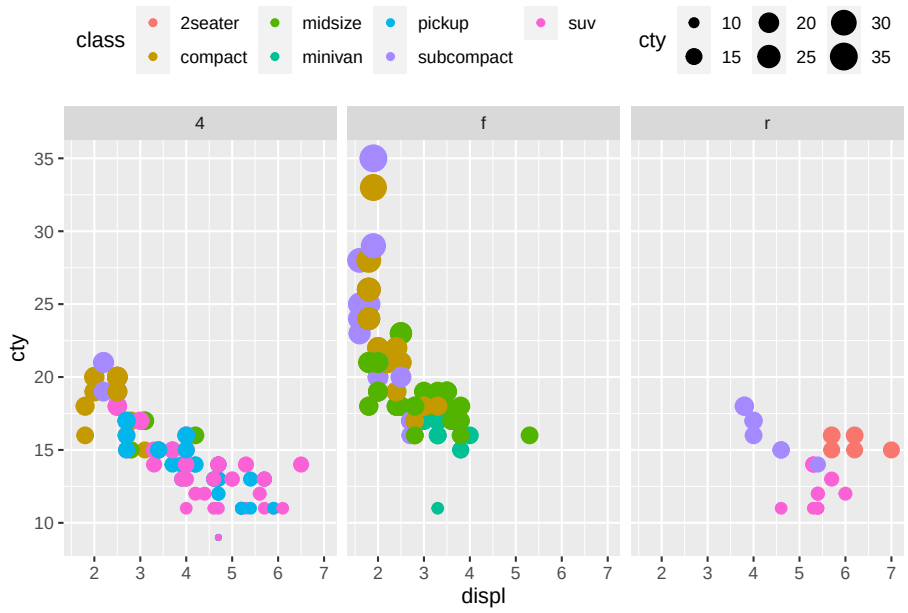
```
esquisser(dados)
```

3.13 Exemplo esquisse

```
ggplot(dados) +
  aes(x = displ, y = hwy, colour = drv) +
  geom_point(shape = "circle", size = 1.85) +
  scale_color_hue(direction = 1) +
  theme_minimal() +
  theme(legend.position = "top")
```




```
ggplot(dados) +
  aes(x = displ, y = cty, colour = class, size = cty) +
  geom_point(shape = "circle") +
  scale_color_hue(direction = 1) +
  theme(legend.position = "top") +
  facet_wrap(vars(drv))
```


class
2seater
compact
midsize
minivan
pickup
subcompact
suv
cty
10
15
20
25
30
35
4
f
r
cty
displ

Capítulo 4

Referências (on-line)

4.1 Livros

ggplot2: elegant graphics for data analysis, Hadley Wickham

R Programming for Data Science, Roger D. Peng

R for Data Science, Hadley Wickham e Garrett Grolemund.

R Graphics Cookbook, Winston Chang

4.2 Tidyverse: links do dplyr, ggplot2 e magrittr

dplyr do tidyverse

ggplot2 do tidyverse

magrittr do tidyverse?

4.3 cheat sheet

Data transformation with dplyr

Data visualization with ggplot2