Recently one of my coworkers showed me a ggplot and although it is not wrong, it is also not ideal. Here is the TL:DR :

Whenever you find yourself adding multiple geom\_\* to show different groups, reshape your data

In software engineering there are things called [antipatterns](https://en.wikipedia.org/wiki/Anti-pattern#Software_engineering), ways of programming  
that lead you into potential trouble. This is one of them.

I’m not saying it is incorrect, but it might lead you into trouble.

example: we have some data, some different calculations and we want to plot that.

**I load tidyverse and create a modified mtcars set in this hidden part,  
but if you don’t care you can leave it unopened**

*Cool how this folds away right? It even works on github markdown, if you want to know how I did this, I explain it* [*here*](https://notes.rmhogervorst.nl/post/2019/02/09/code-folding-github-markdown/)

library(tidyverse) # I started loading magrittr, ggplot2 and tidyr, and realised

## ── Attaching packages ─────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.1.0 ✔ purrr 0.3.0

## ✔ tibble 2.0.1 ✔ dplyr 0.7.8

## ✔ tidyr 0.8.2 ✔ stringr 1.4.0

## ✔ readr 1.3.1 ✔ forcats 0.3.0

## ── Conflicts ────────────────────────────────────────────────── tidyverse\_conflicts() ──

## ✖ dplyr::filter() masks stats::filter()

## ✖ dplyr::lag() masks stats::lag()

# I needed dplyr too, at some point loading tidyverse is simply easiest.

very\_serious\_data <-

mtcars %>%

as\_tibble(rownames = "carname") %>%

group\_by(cyl) %>%

mutate(

mpg\_hp = mpg/hp,

first\_letter = str\_extract(carname, "^[A-z]"),

mpg\_hp\_c = mpg\_hp/mean(mpg\_hp),# grouped mean

mpg\_hp\_am = mpg\_hp+ am

)

Now the data (mtcars) and calculations don’t really make sense but they are here to show you the  
antipattern. I created 3 variants of dividing mpg (miles per gallon) by hp (horse power)

**The antipattern**

We have a dataset with multiple variables (columns) and want to plot  
one against the other, so far so good.

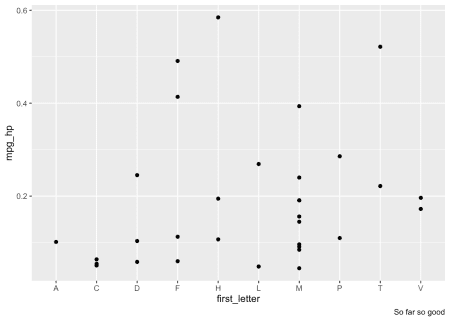
What is the effect of mpg\_hp for every first letter of the cars?

very\_serious\_data %>%

ggplot(aes(first\_letter, mpg\_hp))+

geom\_point()+

labs(caption = "So far so good")



But you might wonder what the other transformations of that variable do?  
You can just add a new geom\_point, but maybe with a different color?  
And to see the dots that overlap you might make them a little opaque.

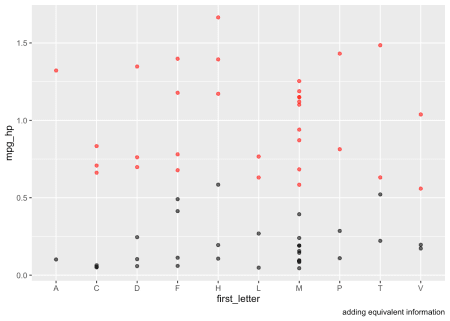
very\_serious\_data %>%

ggplot(aes(first\_letter, mpg\_hp))+

geom\_point(alpha = 2/3)+

geom\_point(aes(y = mpg\_hp\_c), color = "red", alpha = 2/3)+

labs(caption = "adding equivalent information")



And maybe the third one too?

very\_serious\_data %>%

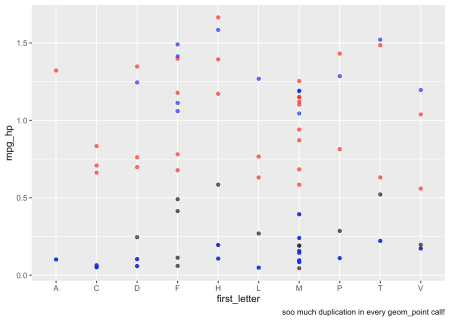
ggplot(aes(first\_letter, mpg\_hp))+

geom\_point(alpha = 2/3)+

geom\_point(aes(y = mpg\_hp\_c), color = "red", alpha = 2/3)+

geom\_point(aes(y = mpg\_hp\_am), color = "blue", alpha = 2/3)+

labs(caption = "soo much duplication in every geom\_point call!")



This results in lots of code duplication for specifying what is essentially  
the same for every geom\_point() call. It’s also really hard to add a legend  
now.

**What is the alternative?**

Whenever you find yourself adding multiple geom\_\* to show different groups, reshape your data

Gather the columns that are essentially representing the group and reshape  
the data into a format more suitable for plotting. Bonus: automatic correct labeling.

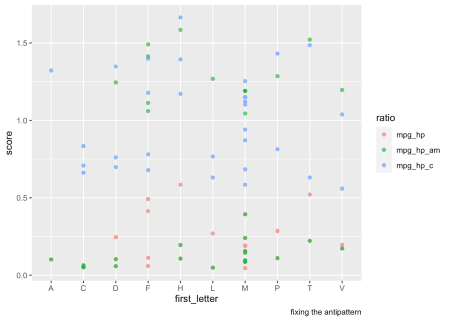
very\_serious\_data %>%

gather(key = "ratio", value = "score", mpg\_hp, mpg\_hp\_c, mpg\_hp\_am ) %>%

ggplot(aes(first\_letter, score, color = ratio))+

geom\_point(alpha = 2/3)+

labs(caption = "fixing the antipattern")



And that’s it.

Mari also tells you it will work

**State of the machine**

At the moment of creation (when I knitted this document ) this was the state of my machine: **click here to expand**

sessioninfo::session\_info()

## ─ Session info ──────────────────────────────────────────────────────────

## setting value

## version R version 3.5.2 (2018-12-20)

## os Ubuntu 16.04.5 LTS

## system x86\_64, linux-gnu

## ui X11

## language en\_US

## collate en\_US.UTF-8

## ctype en\_US.UTF-8

## tz Europe/Amsterdam

## date 2019-03-07

##

## ─ Packages ──────────────────────────────────────────────────────────────

## package \* version date lib source

## assertthat 0.2.0 2017-04-11 [1] CRAN (R 3.5.0)

## backports 1.1.3 2018-12-14 [1] CRAN (R 3.5.2)

## bindr 0.1.1 2018-03-13 [1] CRAN (R 3.5.0)

## bindrcpp \* 0.2.2 2018-03-29 [1] CRAN (R 3.5.0)

## blogdown 0.9 2018-10-23 [1] CRAN (R 3.5.2)

## bookdown 0.9 2018-12-21 [1] CRAN (R 3.5.2)

## broom 0.5.1 2018-12-05 [1] CRAN (R 3.5.2)

## cellranger 1.1.0 2016-07-27 [1] CRAN (R 3.5.0)

## cli 1.0.1 2018-09-25 [1] CRAN (R 3.5.1)

## colorspace 1.4-0 2019-01-13 [1] CRAN (R 3.5.2)

## crayon 1.3.4 2017-09-16 [1] CRAN (R 3.5.0)

## digest 0.6.18 2018-10-10 [1] CRAN (R 3.5.2)

## dplyr \* 0.7.8 2018-11-10 [1] CRAN (R 3.5.1)

## evaluate 0.13 2019-02-12 [1] CRAN (R 3.5.2)

## forcats \* 0.3.0 2018-02-19 [1] CRAN (R 3.5.0)

## generics 0.0.2 2018-11-29 [1] CRAN (R 3.5.2)

## ggplot2 \* 3.1.0 2018-10-25 [1] CRAN (R 3.5.2)

## glue 1.3.0 2018-07-17 [1] CRAN (R 3.5.1)

## gtable 0.2.0 2016-02-26 [1] CRAN (R 3.5.0)

## haven 2.0.0 2018-11-22 [1] CRAN (R 3.5.2)

## hms 0.4.2 2018-03-10 [1] CRAN (R 3.5.0)

## htmltools 0.3.6 2017-04-28 [1] CRAN (R 3.5.0)

## httr 1.4.0 2018-12-11 [1] CRAN (R 3.5.1)

## jsonlite 1.6 2018-12-07 [1] CRAN (R 3.5.1)

## knitr 1.21 2018-12-10 [1] CRAN (R 3.5.2)

## labeling 0.3 2014-08-23 [1] CRAN (R 3.5.0)

## lattice 0.20-38 2018-11-04 [4] CRAN (R 3.5.1)

## lazyeval 0.2.1 2017-10-29 [1] CRAN (R 3.5.0)

## lubridate 1.7.4 2018-04-11 [1] CRAN (R 3.5.0)

## magrittr 1.5 2014-11-22 [1] CRAN (R 3.5.0)

## modelr 0.1.2 2018-05-11 [1] CRAN (R 3.5.0)

## munsell 0.5.0 2018-06-12 [1] CRAN (R 3.5.0)

## nlme 3.1-137 2018-04-07 [4] CRAN (R 3.5.0)

## pillar 1.3.1 2018-12-15 [1] CRAN (R 3.5.2)

## pkgconfig 2.0.2 2018-08-16 [1] CRAN (R 3.5.1)

## plyr 1.8.4 2016-06-08 [1] CRAN (R 3.5.0)

## purrr \* 0.3.0 2019-01-27 [1] CRAN (R 3.5.2)

## R6 2.4.0 2019-02-14 [1] CRAN (R 3.5.2)

## Rcpp 1.0.0 2018-11-07 [1] CRAN (R 3.5.1)

## readr \* 1.3.1 2018-12-21 [1] CRAN (R 3.5.2)

## readxl 1.2.0 2018-12-19 [1] CRAN (R 3.5.2)

## rlang 0.3.1 2019-01-08 [1] CRAN (R 3.5.2)

## rmarkdown 1.11 2018-12-08 [1] CRAN (R 3.5.2)

## rstudioapi 0.8 2018-10-02 [1] CRAN (R 3.5.1)

## rvest 0.3.2 2016-06-17 [1] CRAN (R 3.5.0)

## scales 1.0.0 2018-08-09 [1] CRAN (R 3.5.1)

## sessioninfo 1.1.1 2018-11-05 [1] CRAN (R 3.5.2)

## stringi 1.3.1 2019-02-13 [1] CRAN (R 3.5.2)

## stringr \* 1.4.0 2019-02-10 [1] CRAN (R 3.5.2)

## tibble \* 2.0.1 2019-01-12 [1] CRAN (R 3.5.2)

## tidyr \* 0.8.2 2018-10-28 [1] CRAN (R 3.5.1)

## tidyselect 0.2.5 2018-10-11 [1] CRAN (R 3.5.1)

## tidyverse \* 1.2.1 2017-11-14 [1] CRAN (R 3.5.0)

## withr 2.1.2 2018-03-15 [1] CRAN (R 3.5.0)

## xfun 0.4 2018-10-23 [1] CRAN (R 3.5.2)

## xml2 1.2.0 2018-01-24 [1] CRAN (R 3.5.0)

## yaml 2.2.0 2018-07-25 [1] CRAN (R 3.5.1)

##

## [1] /home/roel/R/x86\_64-pc-linux-gnu-library/3.5

## [2] /usr/local/lib/R/site-library

## [3] /usr/lib/R/site-library

## [4] /usr/lib/R/library