R-Ladies
Amsterdam

GGplot2

By Olga Ivanova, & Sarah Stolle, Janine Khuc

1) Intro and Theory to visualisations
2) Introduction to ggplot
3) Playtime I
4) More advanced ggplot
5) Playtime II

1) Intro and Theory to visualisations
2) Introduction to ggplot
3) Playtime I
4) More advanced ggplot
5) Playtime II

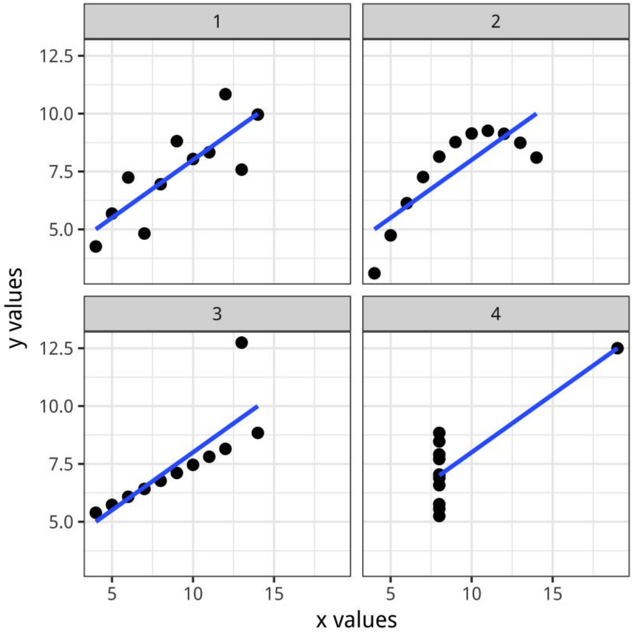
Why do we need data visualization?

Data Exploration

- Find outliers
- See patterns

Visual communication

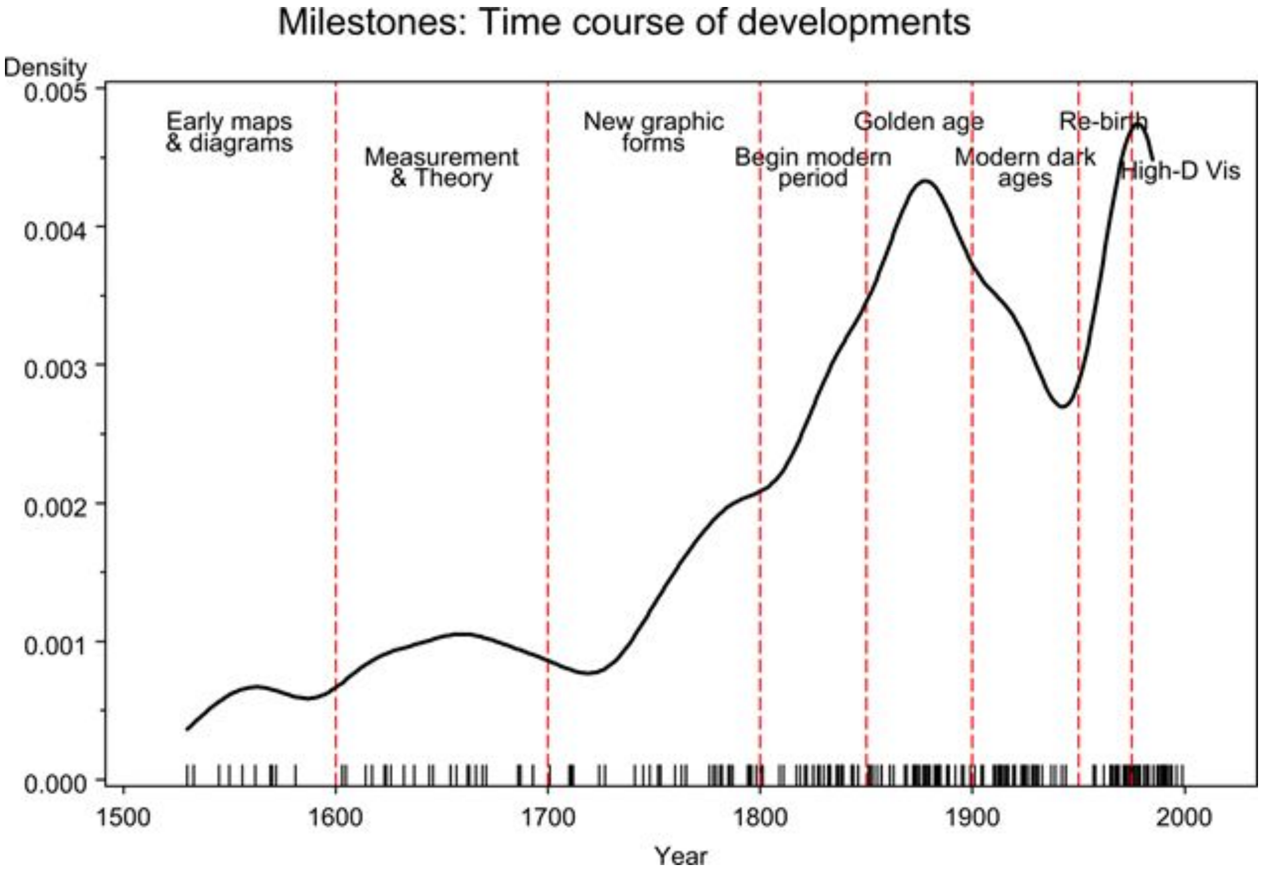
- Convey a message (graph could be clearer than text)
- Summarize data trends



Anscombe's quartet (1973)

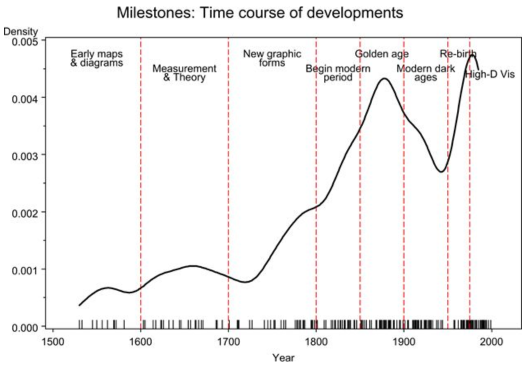
@RLadiesAMS
#RLadies #rstats

A little history of data visualization

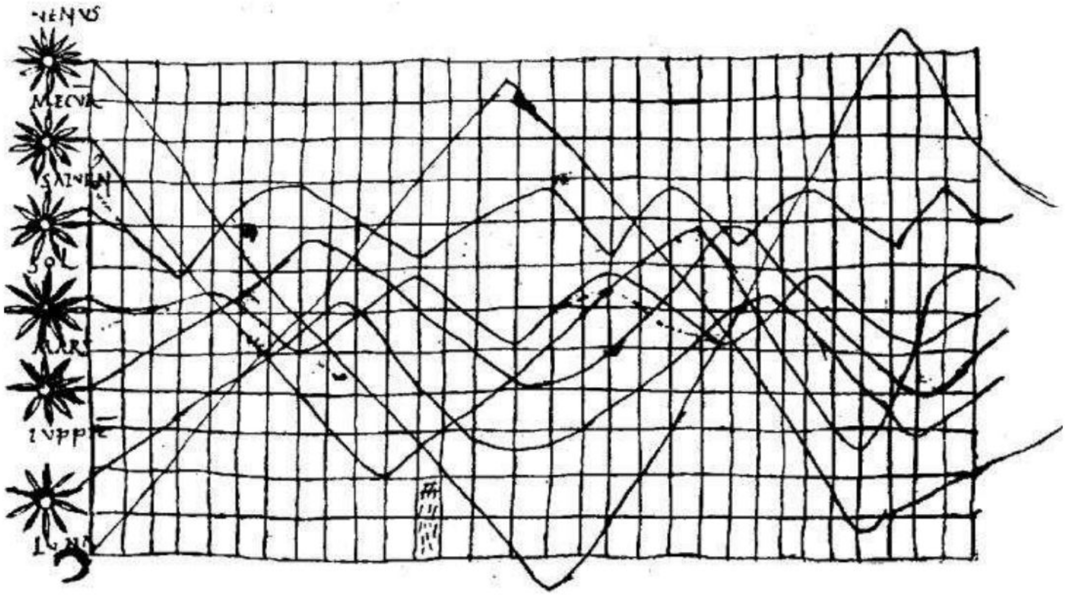


Michael Friendly – A brief history of data visualization (2006)

A little history of data visualization

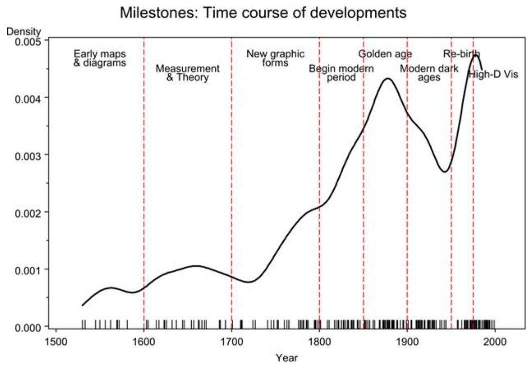


Planetary movements (10th century)



Both graphics: Michael Friendly – A brief history of data visualization (2006)

A little history of data visualization



William Playfair 1821: Time series of prices and wages over 250 years

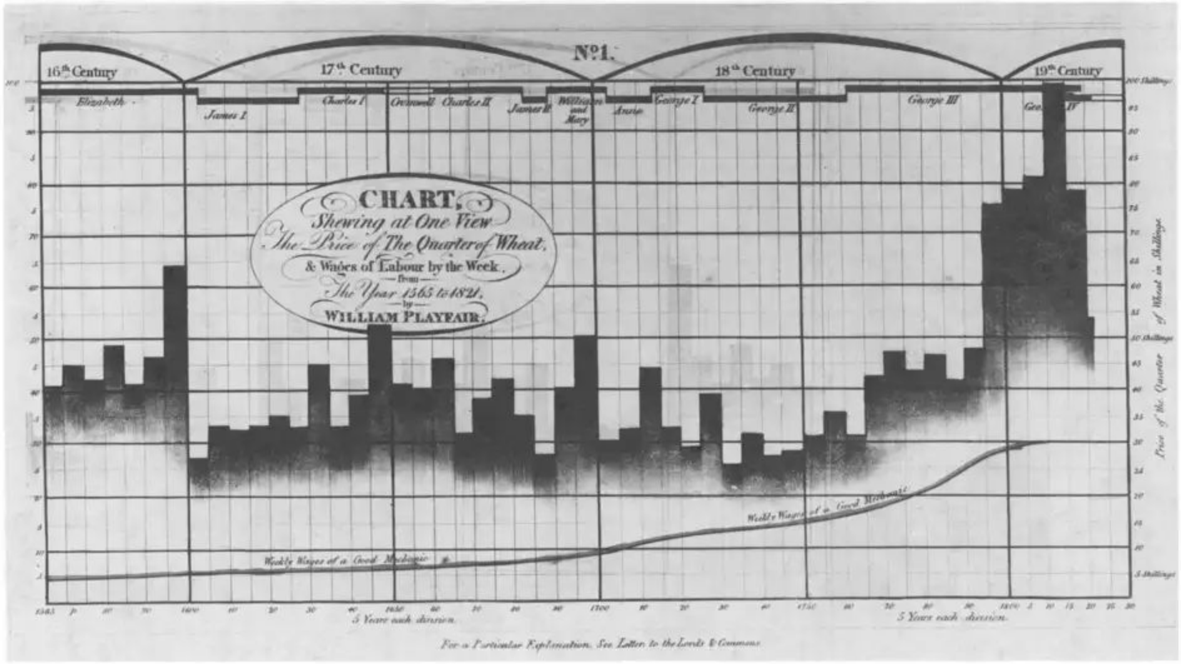


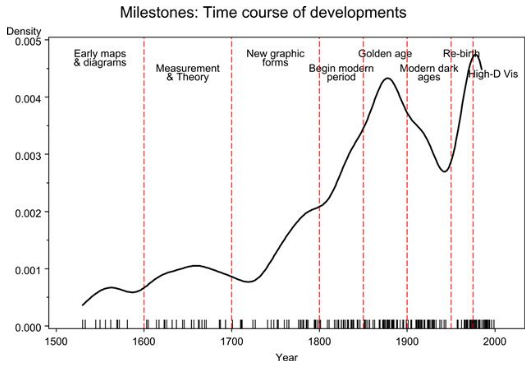
Fig. 3. *One of Playfair's drama graphics designed to show the unfair prices of a quarter of wheat compared with the humble wages of a "good mechanic." In the original, the curve for the wages of the "good mechanic" appears red and the area underneath is stained blue. This chart appears in all three editions of* Agricultural Distresses [*Playfair* (1821, 1822)].

William Playfair

- Line graphs
- Bar graphs
- Pie charts
- Circles

Both graphics: Michael Friendly – A brief history of data visualization (2006)

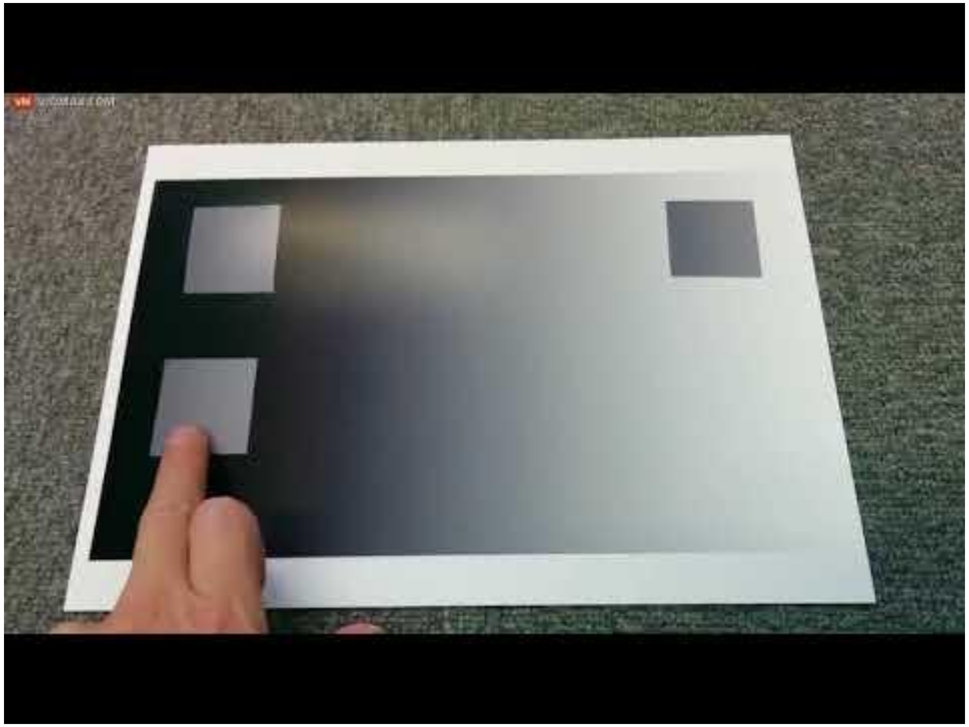
A little history of data visualization



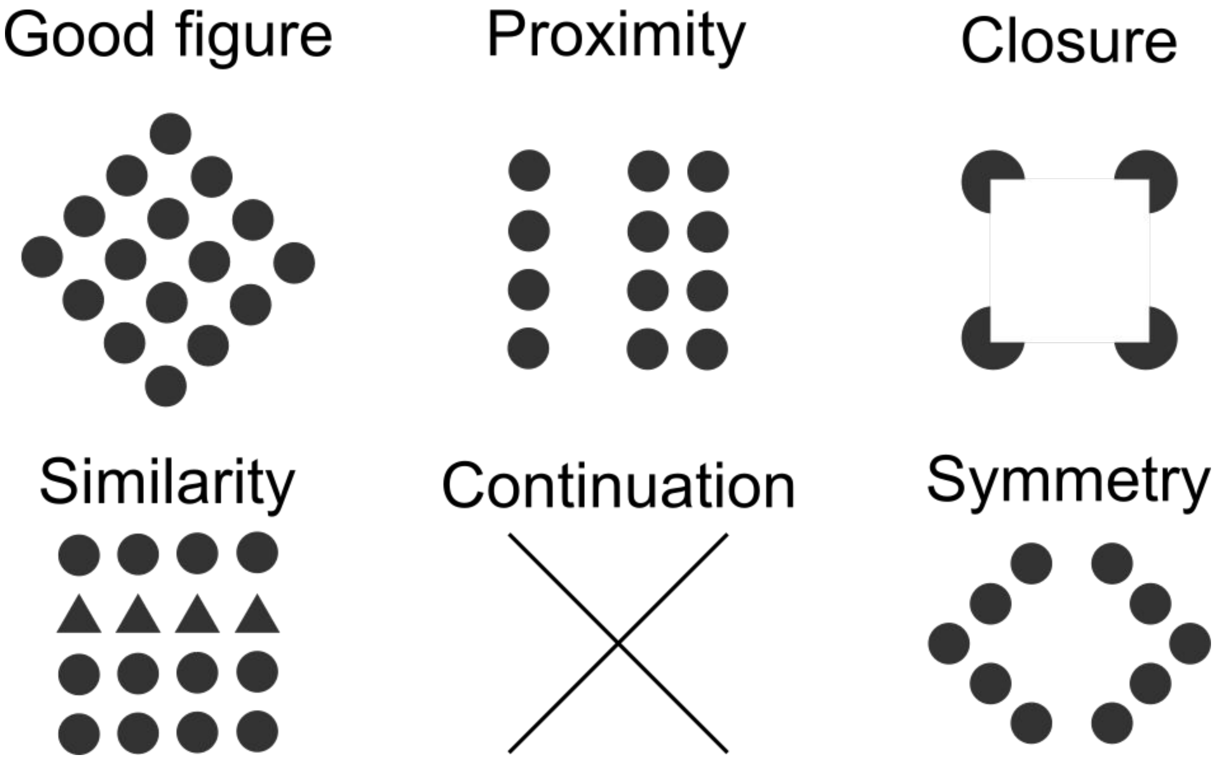
Edward Tufte

Both graphics: Michael Friendly – A brief history of data visualization (2006)

Theory of human visual perception



Perception of patterns - Gestalt principles



inspired by uxcheat.com

Hierarchy of visual perception

Hierarchy of discernibility:

- Colors
- Size
- Alignment
- Character

Question yourself when creating a graph:

- What is your message?

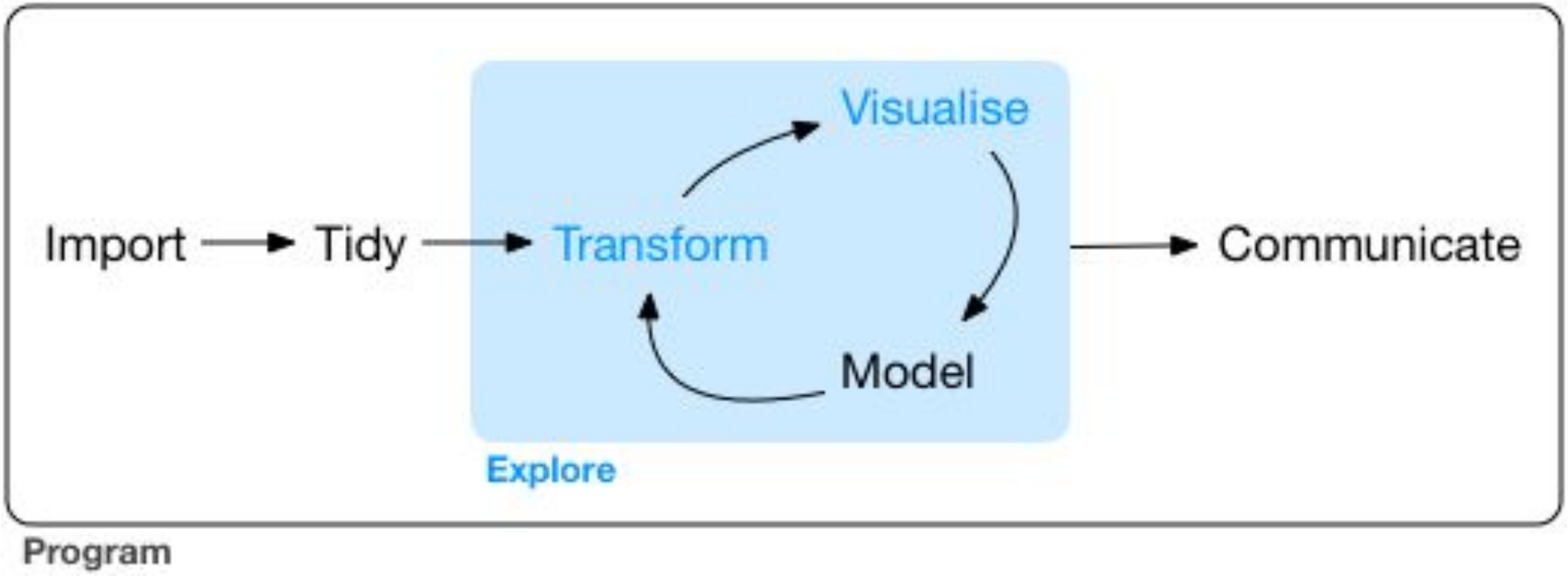
- Who is your audience? (yourself, academic, general public,...)

- Where will it be shown? (article, news, conference, online, paper, …)

R-Ladies
Amsterdam

1) Intro and Theory to visualisations
2) Introduction to ggplot
3) Playtime I
4) More advanced ggplot
5) Playtime II

GGplot in tidyverse



GGplot2

- **GGplot2** (Wickham, 2009) is based on the grammar of graphics (Wilkinson et al., 2006)
    - gg in ggplot means Grammar of Graphics

- Way of thinking about how to create graphs

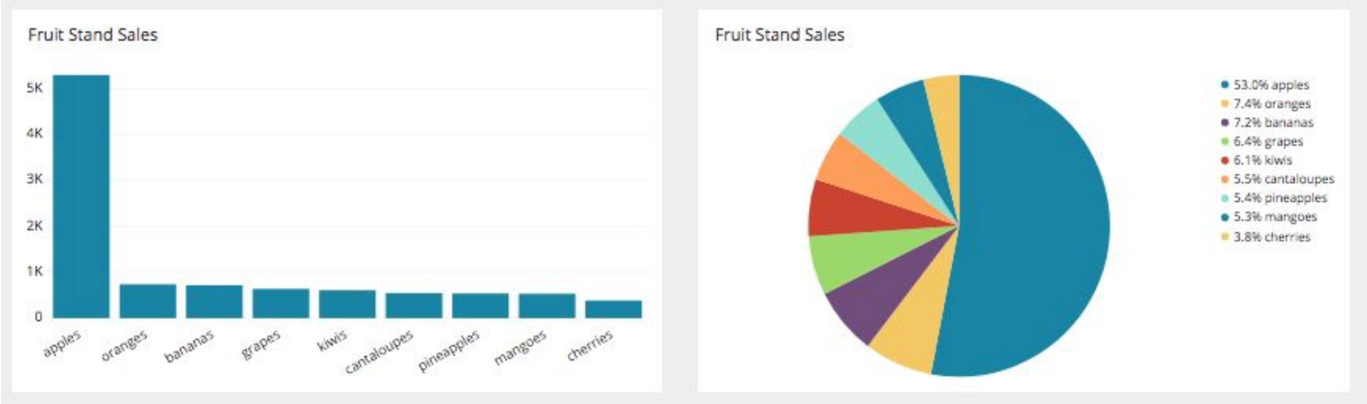
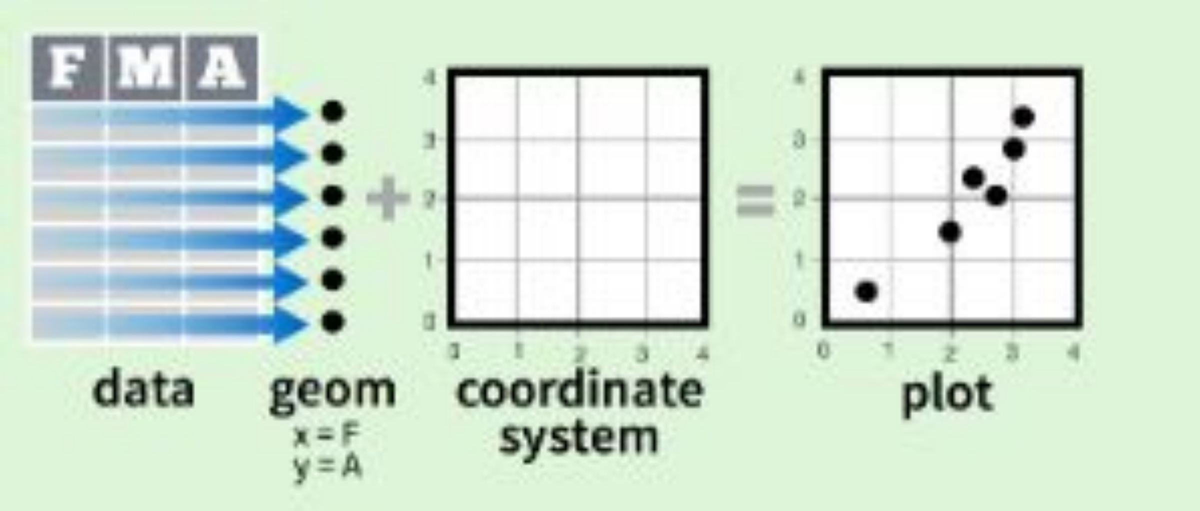


Image from https://chartio.com/learn/dashboards-and-charts/what-is-the-difference-between-a-pie-chart-and-a-bar-chart/

GGplot2

- **data**set, a **coordinate system** and **geoms,** mapping visual properties to geom (**aesthetics**) - e.g. **size, color**



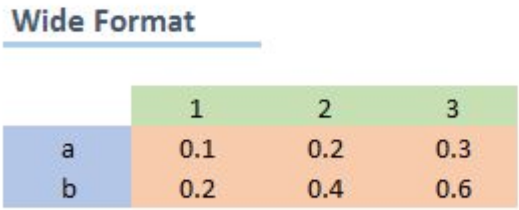
From ggplot cheatsheet

GGplot input

- Uses data frames as input
  - **Long format**

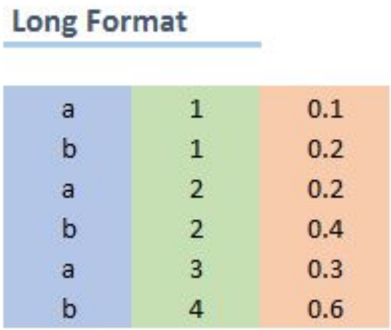
    Each row - observation of a particular category

  - Typically use `reshape2` or `tidyr` to get data in long format



Wide Format

	1	2	3
a	0.1	0.2	0.3
b	0.2	0.4	0.6



Long Format

a	1	0.1
b	1	0.2
a	2	0.2
b	2	0.4
a	3	0.3
b	4	0.6

Layers

```
library(ggplot2)
… # load some data
ggplot(fastfood_calories) + …
```

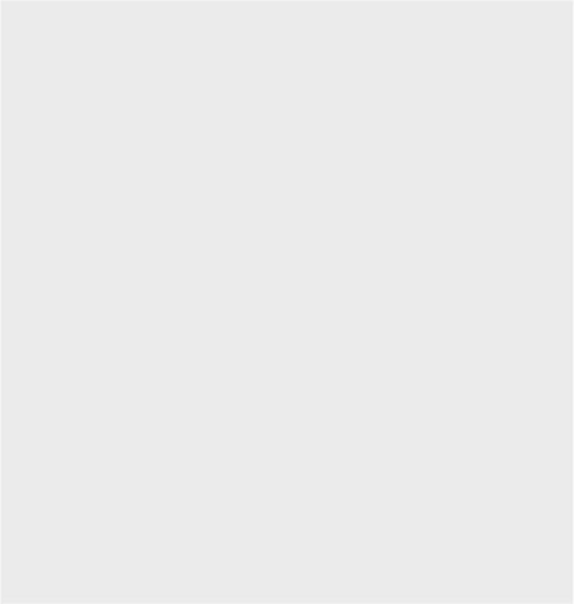


Image from https://www.shutterstock.com/image-photo/fresh-baked-spekkoek-traditional-indonesian-layer-1104518450?src=Y3gY2z6y-Is7QbDAjV1h2g-1-1

aes (aesthetics)

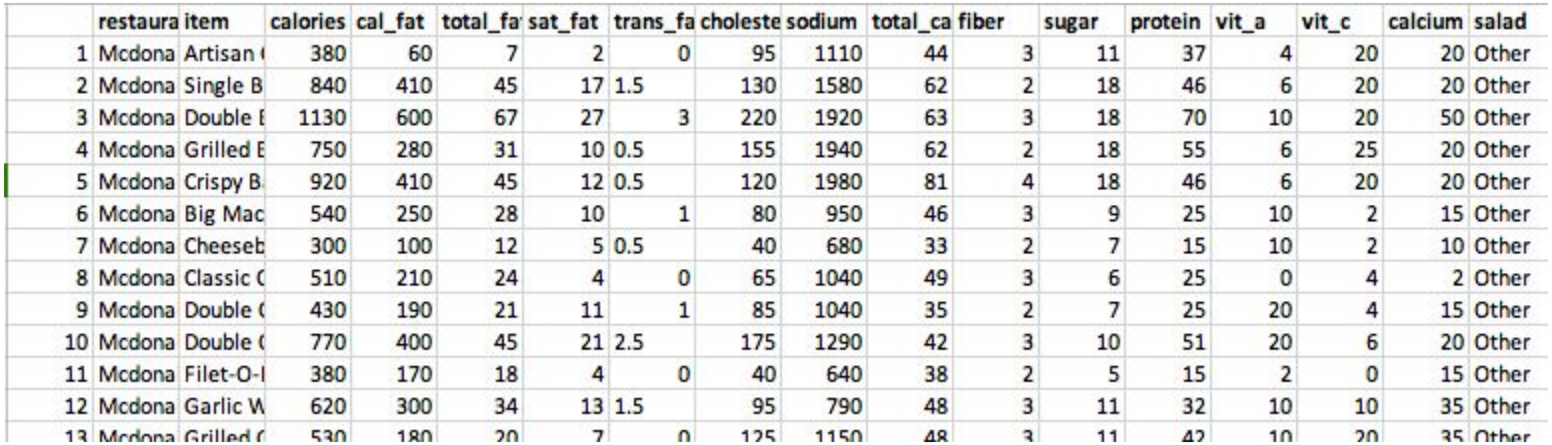
- **Aesthetics**
  - Horizontal
  - Vertical
  - Color
  - Different lines
  - Line type
  - Size
  - Shape
  - ….

- In `ggplot2` we first set aesthetic mapping of our data using `aes()` inside `ggplot()`

```
library(ggplot2)
ggplot(fastfood_calories, aes(x=calories,
y=total_fat))
```

Data introduction

Original data come from fastfoodnutrition.org, dataset with nutritional information about entrees (main courses) from the specified fast food franchises

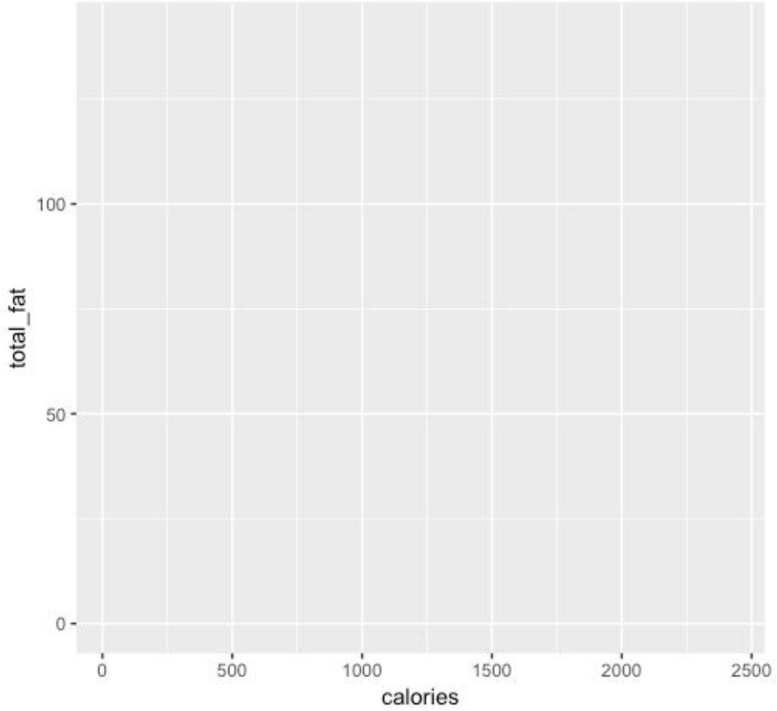


	restaura	item	calories	cal_fat	total_fa	sat_fat	trans_fa	choleste	sodium	total_ca	fiber	sugar	protein	vit_a	vit_c	calcium	salad
1	Mcdona	Artisan (	380	60	7	2	0	95	1110	44	3	11	37	4	20	20	Other
2	Mcdona	Single B	840	410	45	17	1.5	130	1580	62	2	18	46	6	20	20	Other
3	Mcdona	Double E	1130	600	67	27	3	220	1920	63	3	18	70	10	20	50	Other
4	Mcdona	Grilled E	750	280	31	10	0.5	155	1940	62	2	18	55	6	25	20	Other
5	Mcdona	Crispy B.	920	410	45	12	0.5	120	1980	81	4	18	46	6	20	20	Other
6	Mcdona	Big Mac	540	250	28	10	1	80	950	46	3	9	25	10	2	15	Other
7	Mcdona	Cheeseb	300	100	12	5	0.5	40	680	33	2	7	15	10	2	10	Other
8	Mcdona	Classic (	510	210	24	4	0	65	1040	49	3	6	25	0	4	2	Other
9	Mcdona	Double (	430	190	21	11	1	85	1040	35	2	7	25	20	4	15	Other
10	Mcdona	Double (	770	400	45	21	2.5	175	1290	42	3	10	51	20	6	20	Other
11	Mcdona	Filet-O-I	380	170	18	4	0	40	640	38	2	5	15	2	0	15	Other
12	Mcdona	Garlic W	620	300	34	13	1.5	95	790	48	3	11	32	10	10	35	Other
13	Mcdona	Grilled (	530	180	20	7	0	125	1150	48	3	11	42	10	20	35	Other

```
fastfood_calories = read.csv(paste0(folder_path,
"fastfood_calories.csv"), stringsAsFactors = F)
```

aes

```
library(ggplot2)
ggplot(fastfood_calories, aes(x=calories,
y=total_fat))
```



geom (geometics)

```
library(ggplot2)
ggplot(fastfood_calories, aes(x=calories, y=total_fat)) +
geom_point()
```

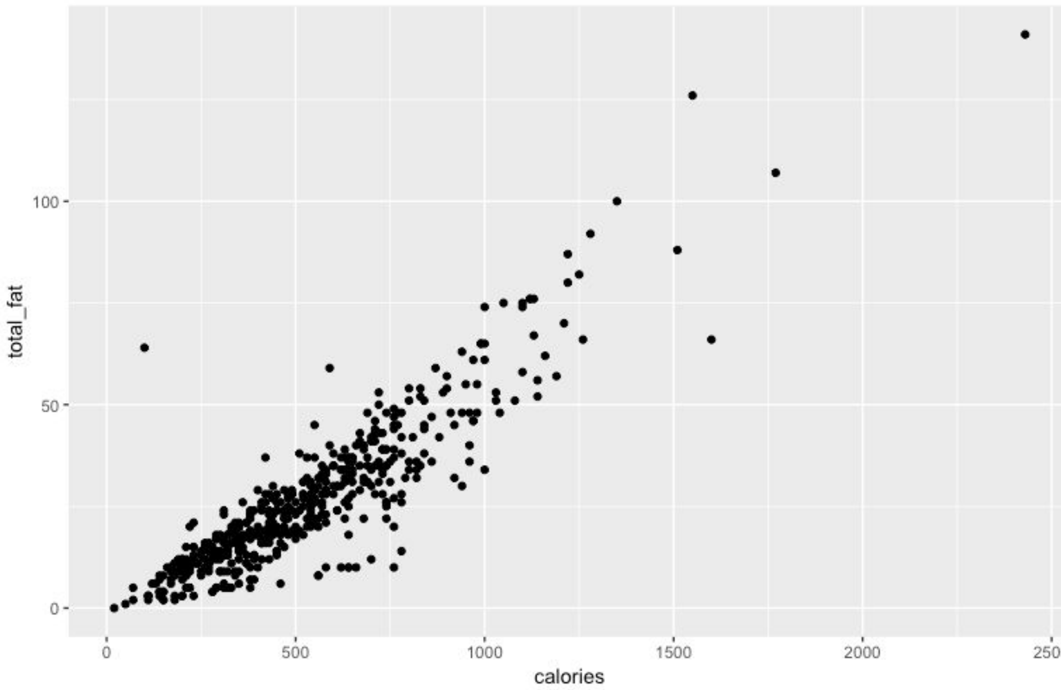
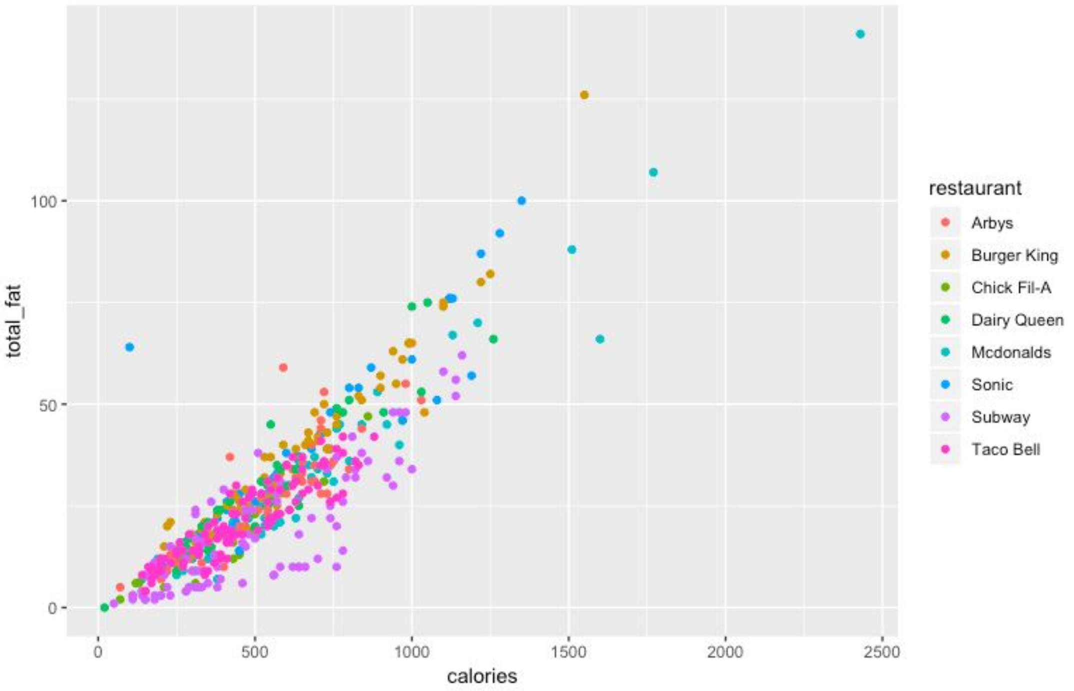


Image from https://pngtree.com/freepng/vector-hamburgers_2978361.html

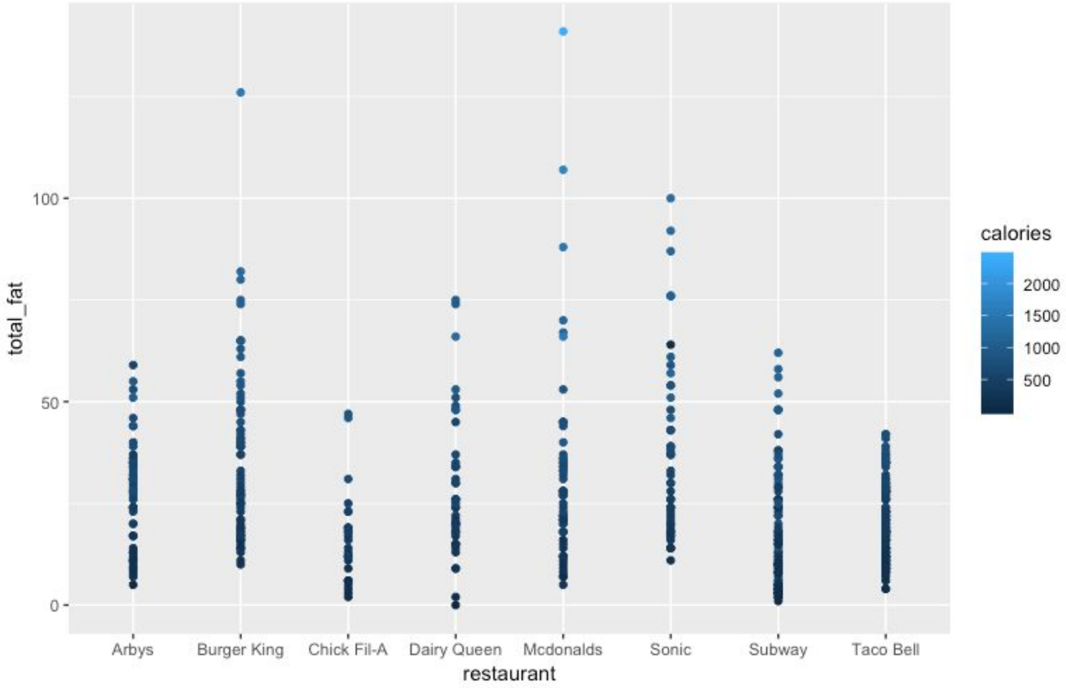
geom

```
library(ggplot2)
ggplot(fastfood_calories, aes(x=calories, y=total_fat,
color=restaurant)) +
geom_point()
```



geom

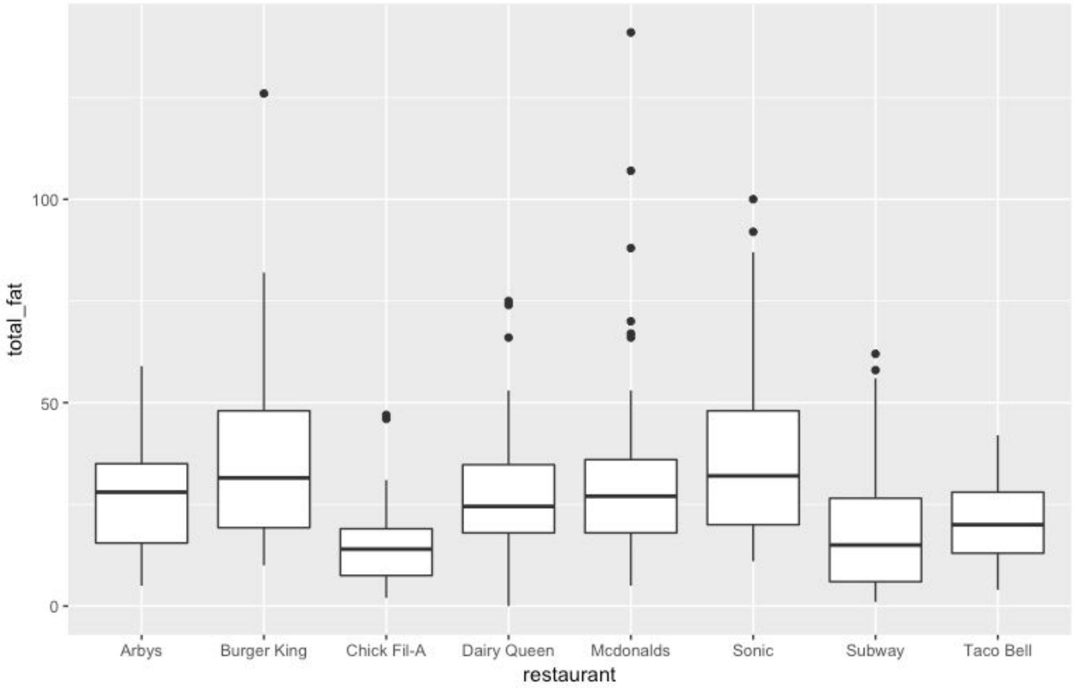
```
library(ggplot2)
ggplot(fastfood_calories, aes(x=restaurant y=total_fat,
color=calories)) +
geom_point()
```



@RLadiesAMS
#RLadies #rstats

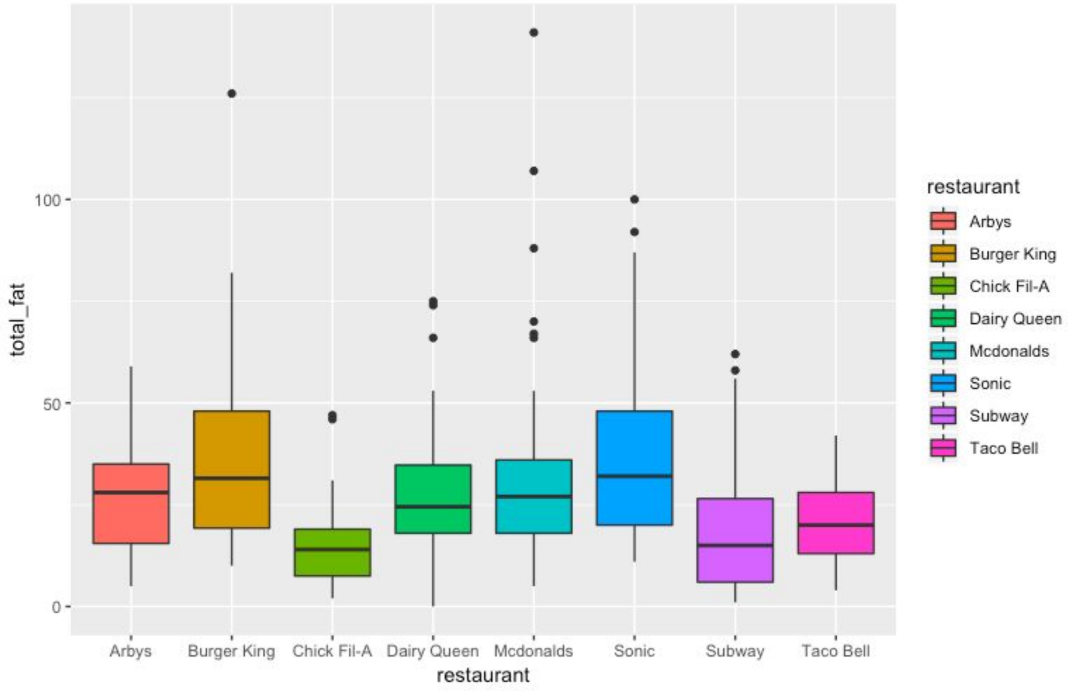
geom: boxplot

```
library(ggplot2)
ggplot(fastfood_calories, aes(x=restaurant y=total_fat)) +
geom_boxplot()
```



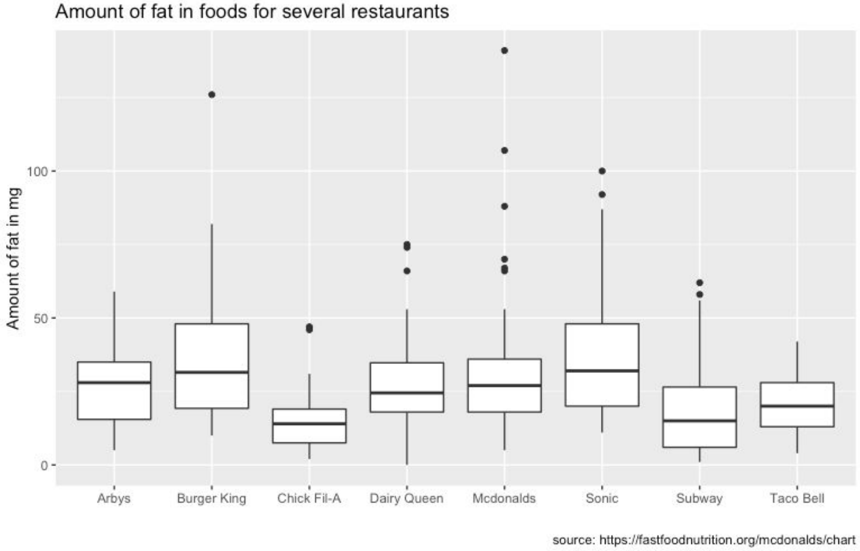
geom: boxplot

```
library(ggplot2)
ggplot(fastfood_calories, aes(x=restaurant y=total_fat,
fill=restaurant)) +
geom_boxplot()
```



Adding labels

```
ggplot(fastfood_calories, aes(x=restaurant y=total_fat)) +
geom_boxplot() +
labs(title= 'Amount of fat in foods for several restaurants',
     x= ' ',  y= 'Amount of fat in mg',
     caption=
'source:https://fastfoodnutrition.org/mcdonalds/chart)
```

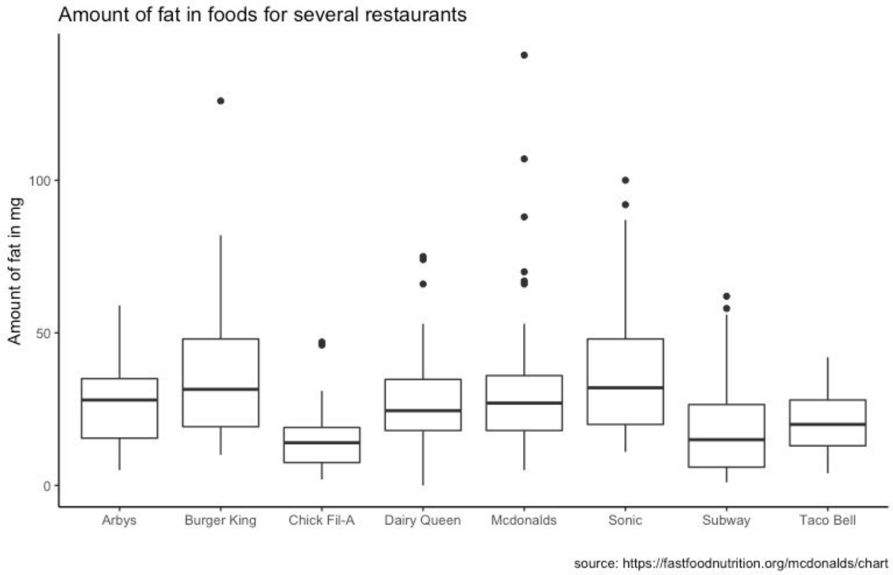


Amount of fat in foods for several restaurants

source: https://fastfoodnutrition.org/mcdonalds/chart

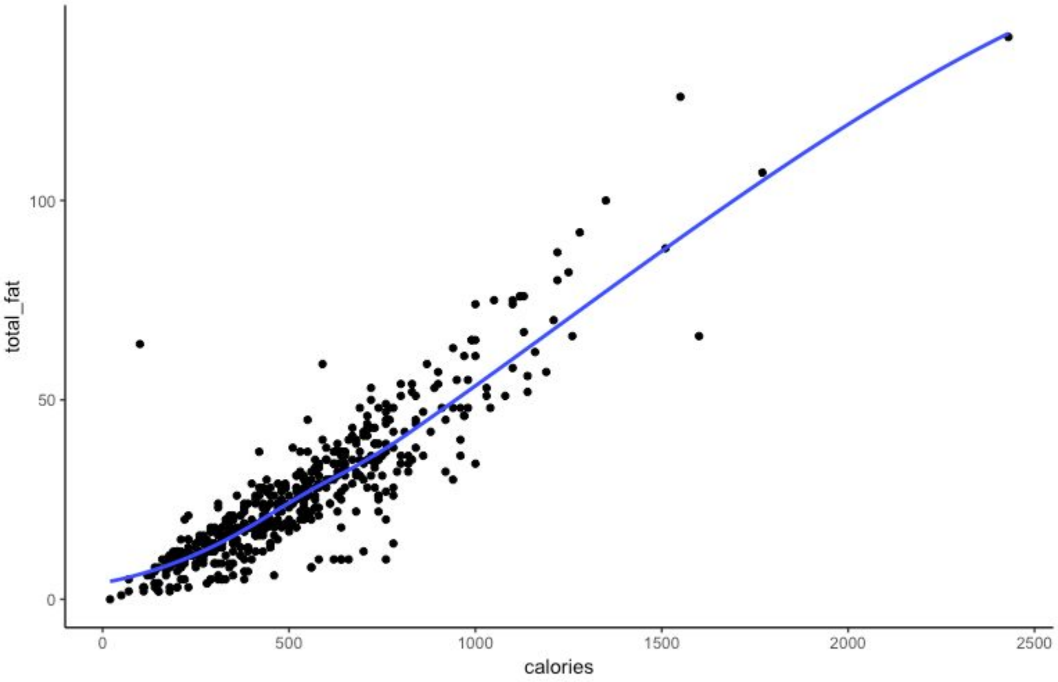
Adding themes

```
ggplot(fastfood_calories, aes(x=restaurant y=total_fat)) +
geom_boxplot() +
labs(title= 'Amount of fat in foods for several restaurants',
     x= ' ',  y= 'Amount of fat in mg',
     caption=
'source:https://fastfoodnutrition.org/mcdonalds/chart) +
theme_classic()
```



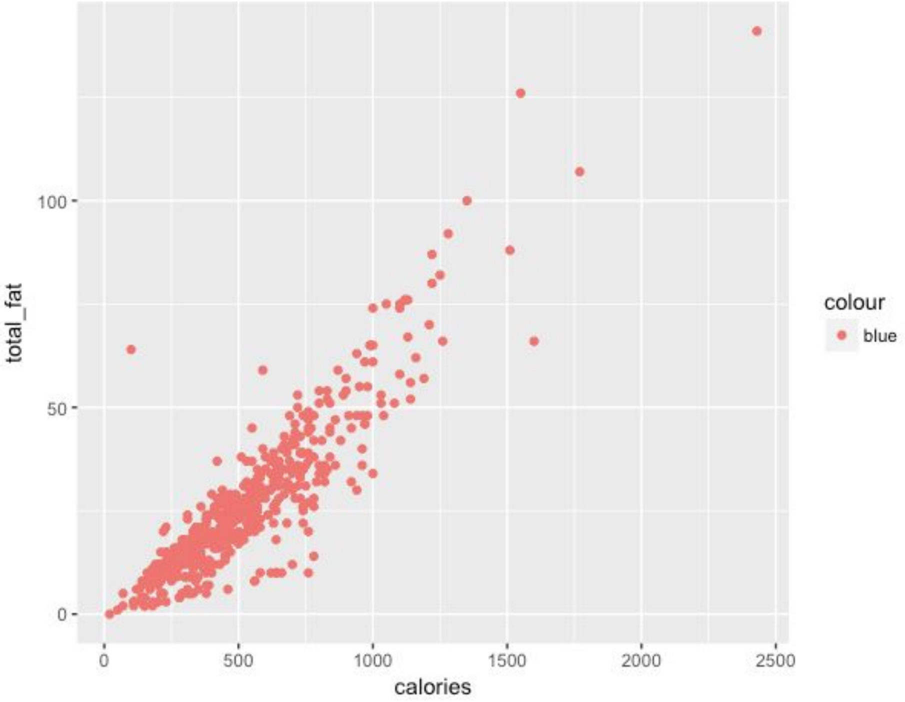
Several geoms

```
ggplot(fastfood_calories, aes(x=calories,
y=total_fat)) +
geom_point()+
geom_smooth(se=FALSE)+
theme_classic()
```



geom and aes: color

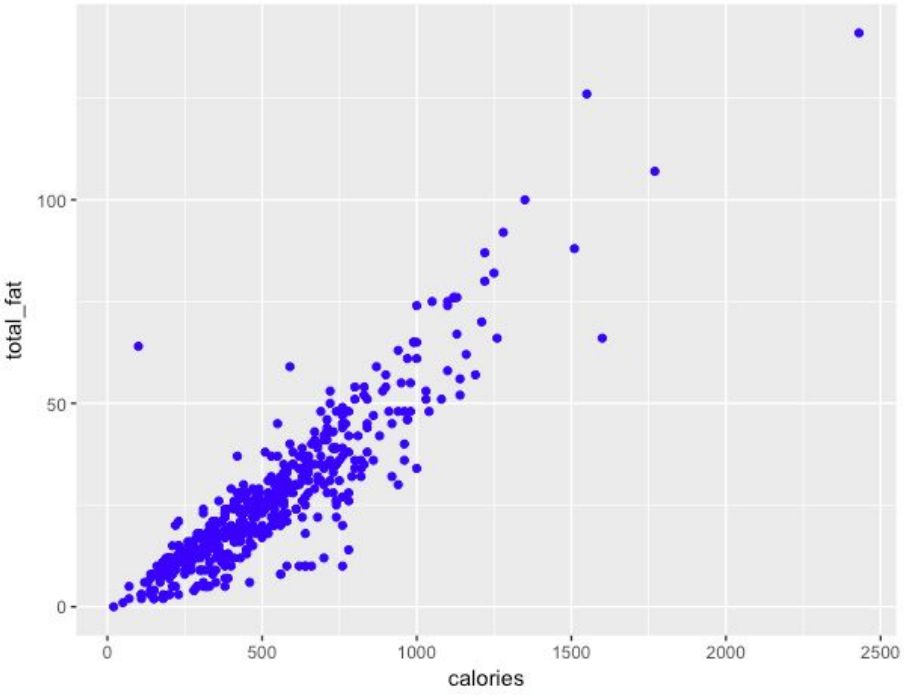
```
library(ggplot2)
ggplot(fastfood_calories, aes(x=calories, y=total_fat,
color="blue")) +
geom_point()
```



@RLadiesAMS
#RLadies #rstats

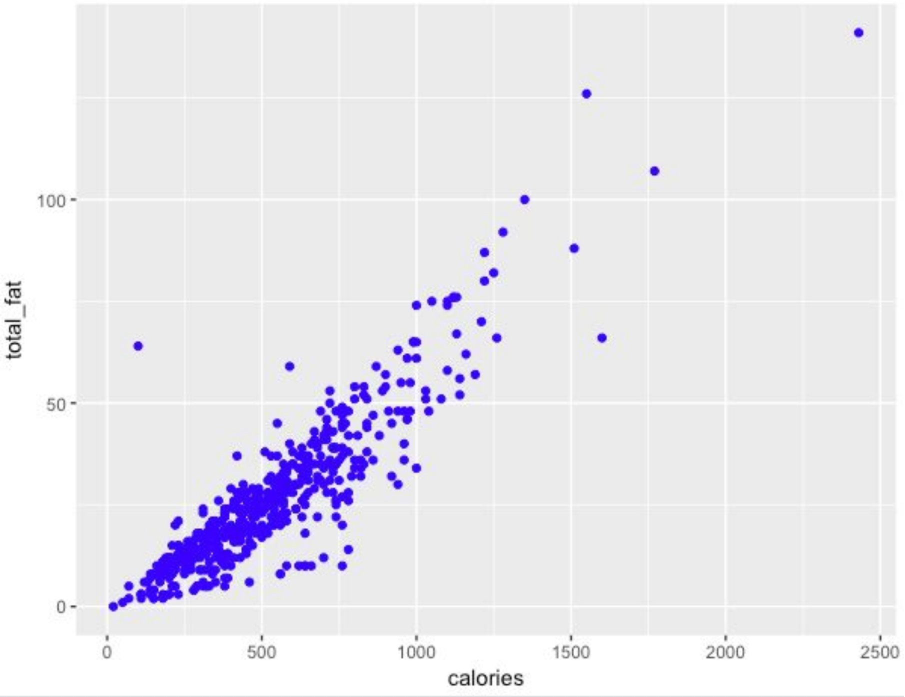
geom and aes: color

```
library(ggplot2)
ggplot(fastfood_calories, aes(x=calories, y=total_fat)) +
geom_point(color="blue")
```



Order is important

```
library(ggplot2)
ggplot(fastfood_calories, aes(x=calories,
y=total_fat, color=restaurant)) +
geom_point(color="blue")
```



geom forms

```
geom_histogram()
geom_point()
geom_smooth()
geom_bar()
```

```
geom_boxplot()
geom_area()
geom_line()
```

Ggplot cheat sheet:
https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf

GGplot reference:
https://ggplot2.tidyverse.org/reference/

Summary

MVP - minimum viable plot with

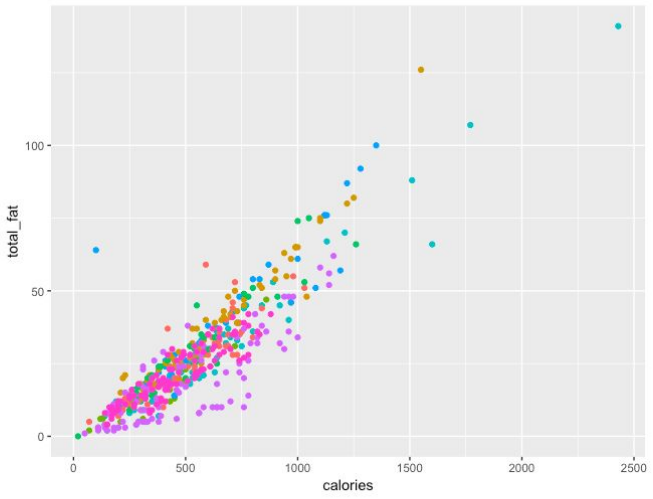
- data
- geom
- aes

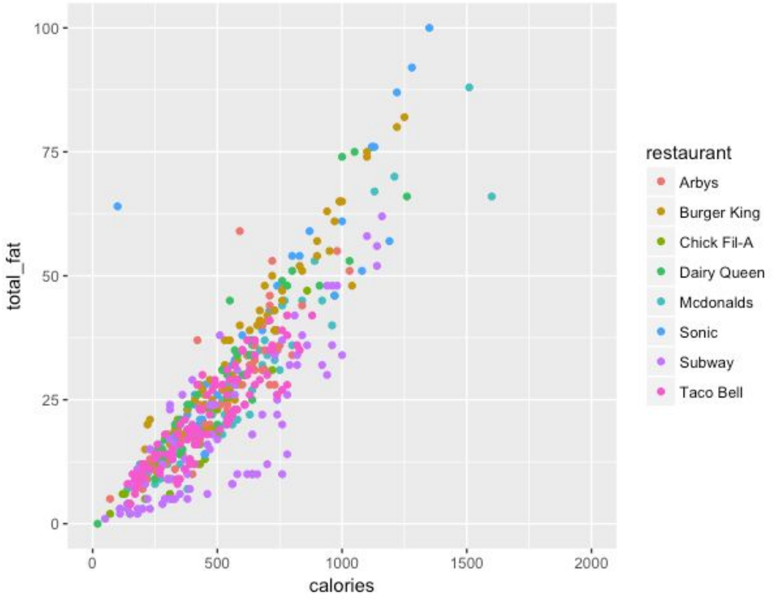
1) Intro and Theory to visualisations
2) Introduction to ggplot
3) Playtime I (15 mins)
4) More advanced ggplot
5) Playtime II

1) Intro and Theory to visualisations
2) Introduction to ggplot
3) Playtime I
4) More advanced ggplot
5) Playtime II

Adjusting axis limits

```
ggplot(fastfood_calories, aes(x=restaurant y=total_fat)) +
geom_point() +
xlim(0, 2000) +
ylim(0, 100)
```



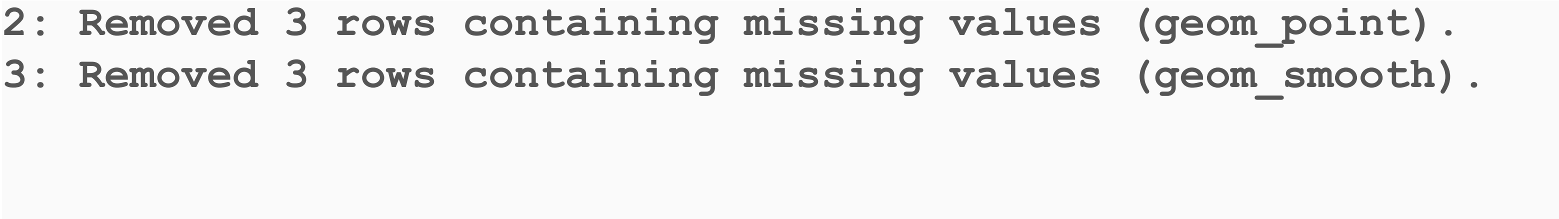


Adjusting axis limits

- `xlim, ylim`

BUT

- `coord_cartesian (xlim, ylim)`

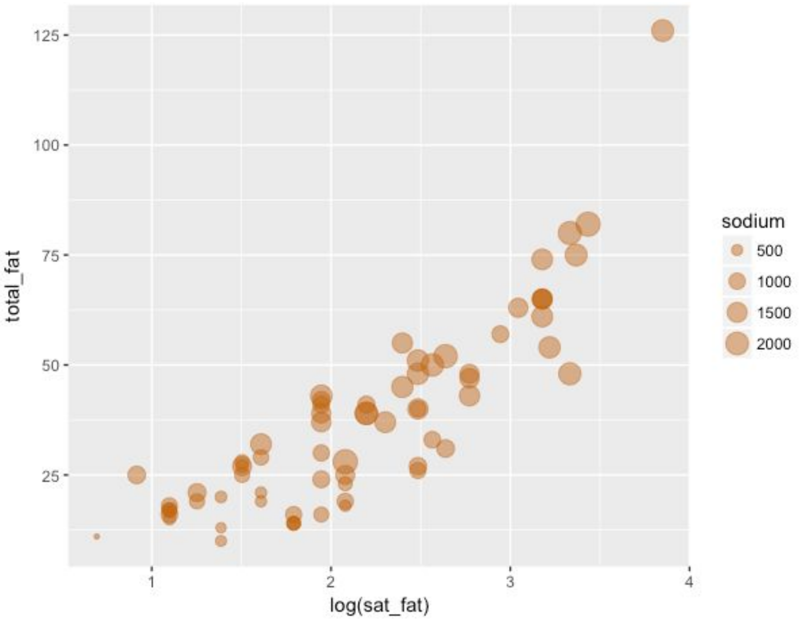


```
2: Removed 3 rows containing missing values (geom_point).
3: Removed 3 rows containing missing values (geom_smooth).
```

Changing points

```
burger_king_rest =
fastfood_calories[fastfood_calories$restaurant=="Burger
King", ]

ggplot(burger_king_rest, aes(x=log(sat_fat), y=total_fat,
size=sodium)) +
geom_point(alpha=0.5, color="#CC6600")
```

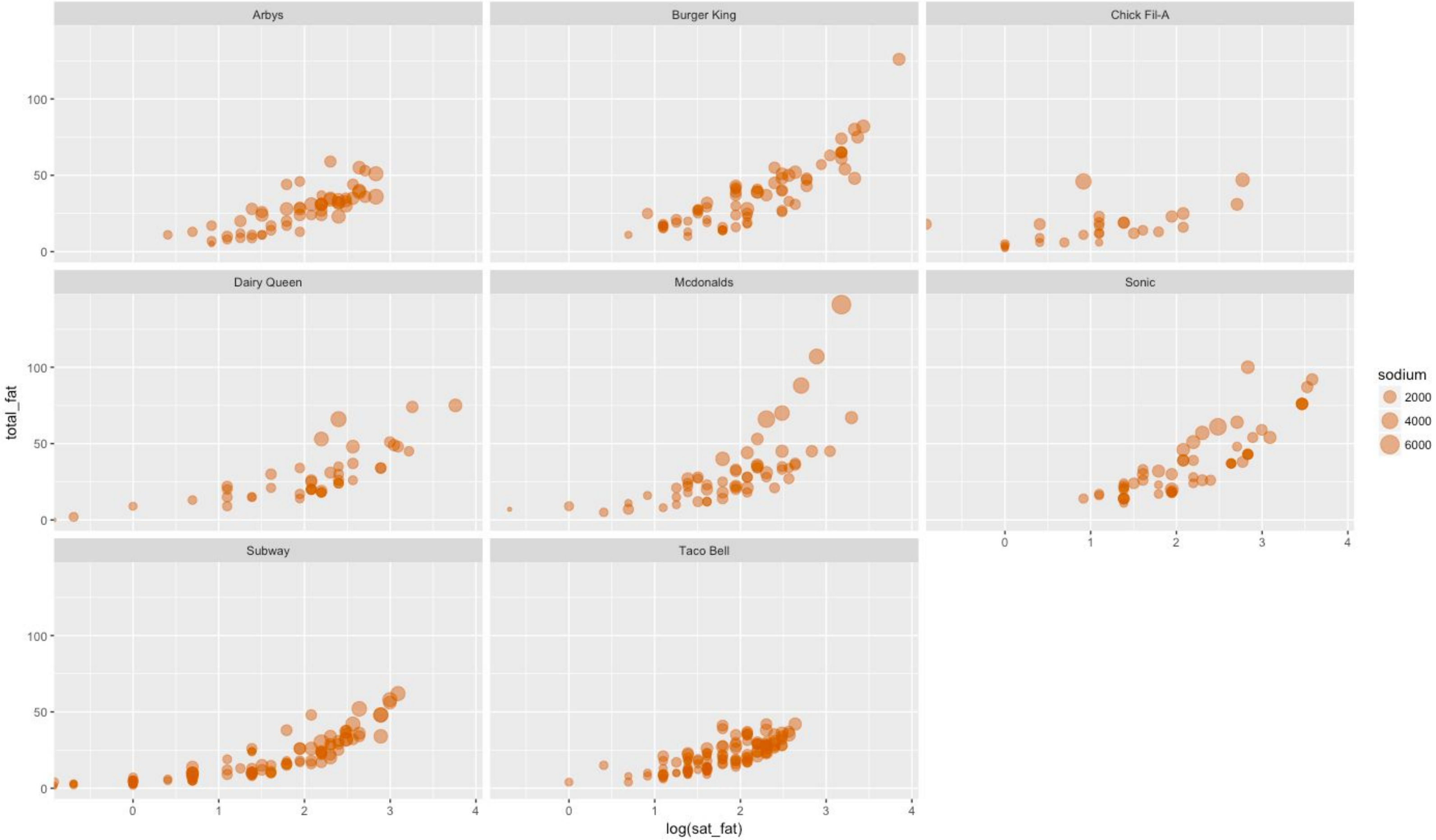


Facets

```
ggplot(fastfood_calories, aes(x=log(sat_fat), y=total_fat,
size=sodium)) +
geom_point(alpha=0.5, color="#CC6600") +
facet_wrap(~restaurant)
```

- facet_wrap or facet_grid

Facets



Arbys
Burger King
Chick Fil-A
Dairy Queen
Mcdonalds
Sonic
Subway
Taco Bell
total_fat
0
50
100
log(sat_fat)
0
1
2
3
4
sodium
2000
4000
6000

Adjusting positions

Minor tweaks to the position of elements, primarily apply to **bars**

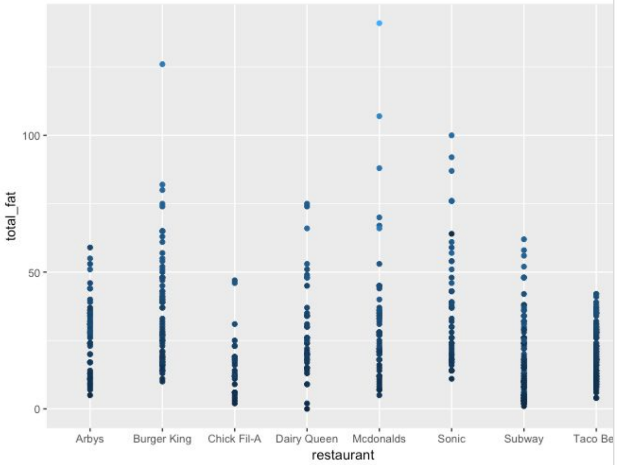
- `position_dodge()`
- `position_stack()`
- `position_fill()`

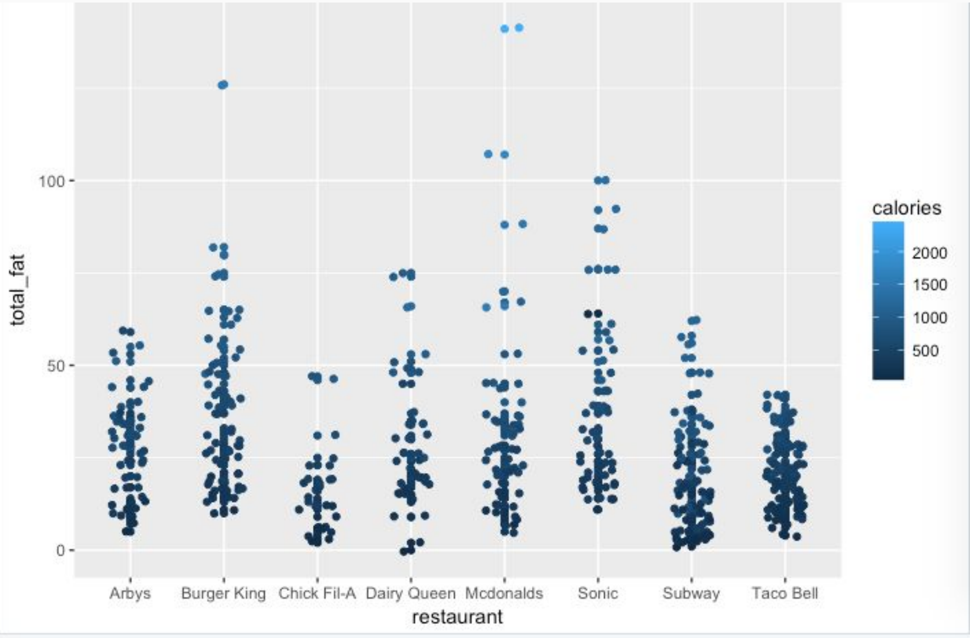
Adjustments for **points**

- `position_judge()`
- `position_jitter()`
- `position_jitterdodge()`

Adjusting positions: jitter

```
ggplot(restr_food_nutrition, aes(x=restaurant, y=total_fat,
color=calories)) +
geom_point() +
geom_jitter(width = 0.2)
```



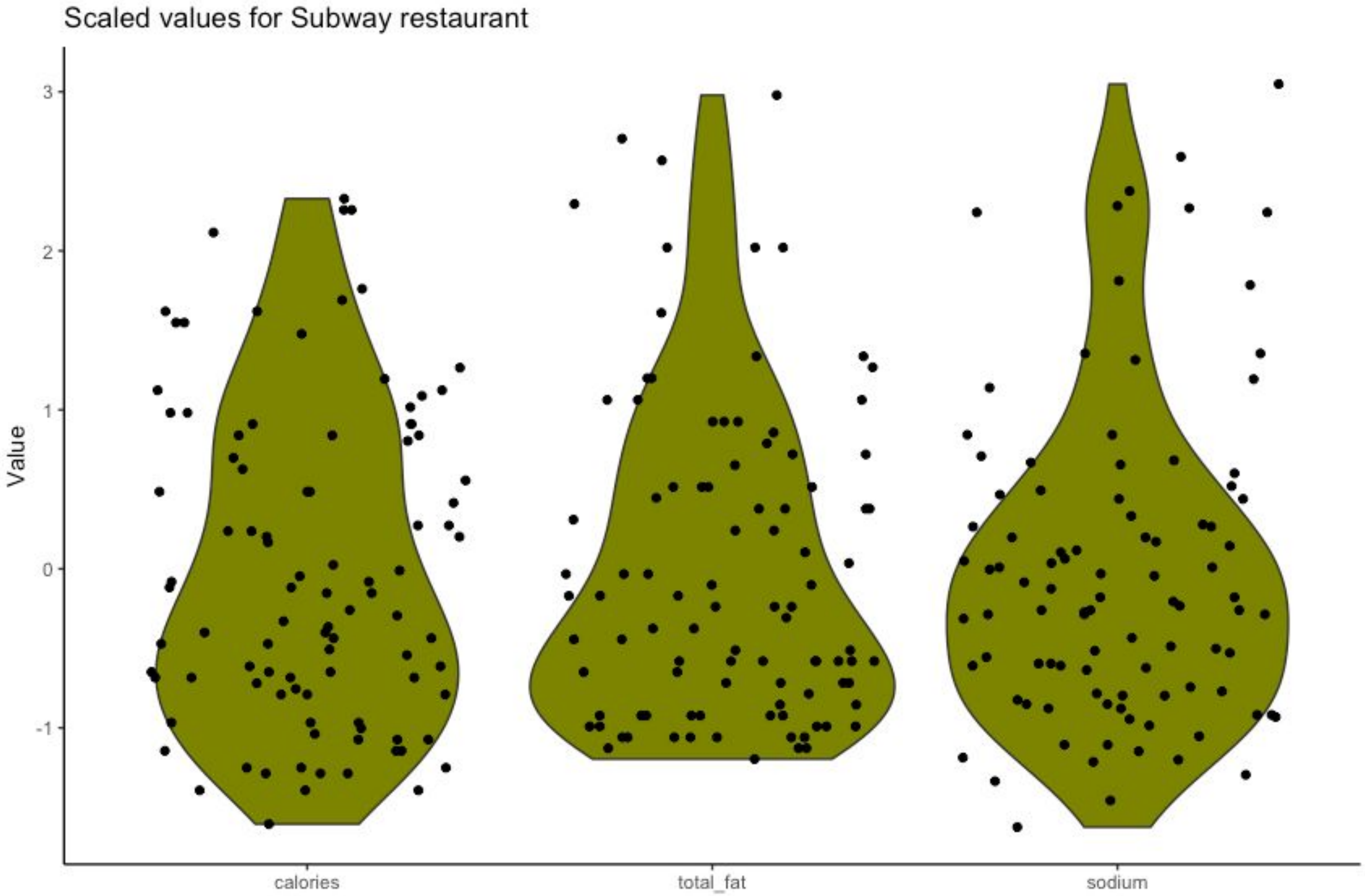


1) Intro and Theory to visualisations
2) Introduction to ggplot
3) Playtime I
4) More advanced ggplot
5) Playtime II

Tidytuesday

- Social data project
- Data released every Monday on https://github.com/rfordatascience/tidytuesday
- Around tidying and visualising the data
- work published on Twitter under #TidyTuesday

Reproduce this plot (exercise 8)



Read further & Inspiration

- Ggplot2 manual
- Data visualization- A practical introduction by Kieran Healy
- TidyTuesday
- Reddit - Data is beautiful
- List with resources about design and code by Lena Groeger
- Data to Vis
- R- graphics
- Fundamentals of data Visualisations

Extra: coordinate systems and saving

- `coord_polar` (to plot Piecharts)
- `coord_map` (to overlay data on geo map)
- `ggsave` (to save your plots)

- Write your own functions like geom or layer!