

# Two principled approaches to data visualization

Stat Bytes

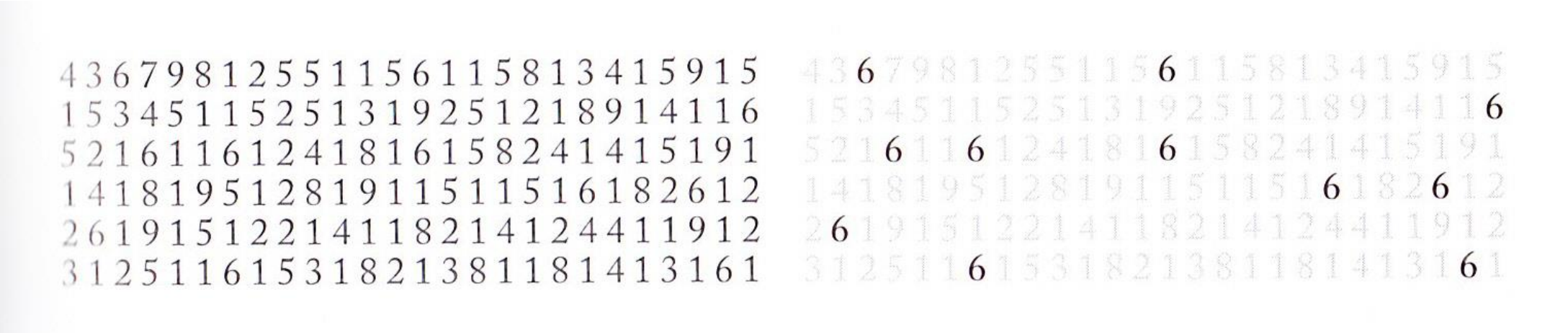
Jerzy Wieczorek

7/8/2015

# How many 6s are there?

43679812551156115813415915  
15345115251319251218914116  
52161161241816158241415191  
14181951281911511516182612  
26191512214118214124411912  
31251161531821381181413161

# Preattentive processing



43679812551156115813415915  
15345115251319251218914116  
52161161241816158241415191  
14181951281911511516182612  
26191512214118214124411912  
31251161531821381181413161

43679812551156115813415915  
15345115251319251218914116  
52161161241816158241415191  
14181951281911511516182612  
26191512214118214124411912  
31251161531821381181413161

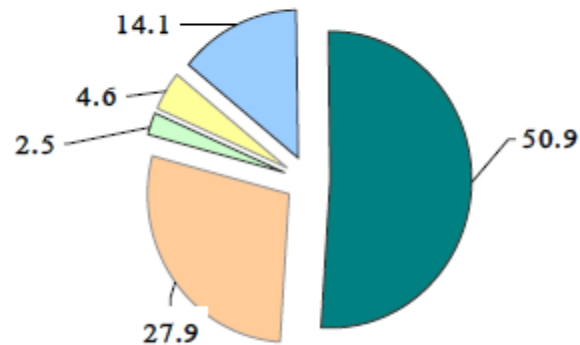
(Alberto Cairo, *The Functional Art*)

# What could be improved here?

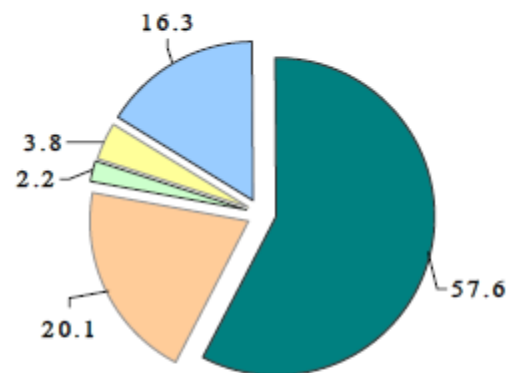
## TOTAL VOLUME OF RETAIL TRADE TURNOVER BY FORMATION SOURCES ОБЩИЙ ОБЪЕМ РОЗНИЧНОГО ТОВАРООБОРОТА ПО ИСТОЧНИКАМ ФОРМИРОВАНИЯ

տոկոսներով՝ ընդհանուրի նկատմամբ - in percent to total - в процентах к итогу

2007



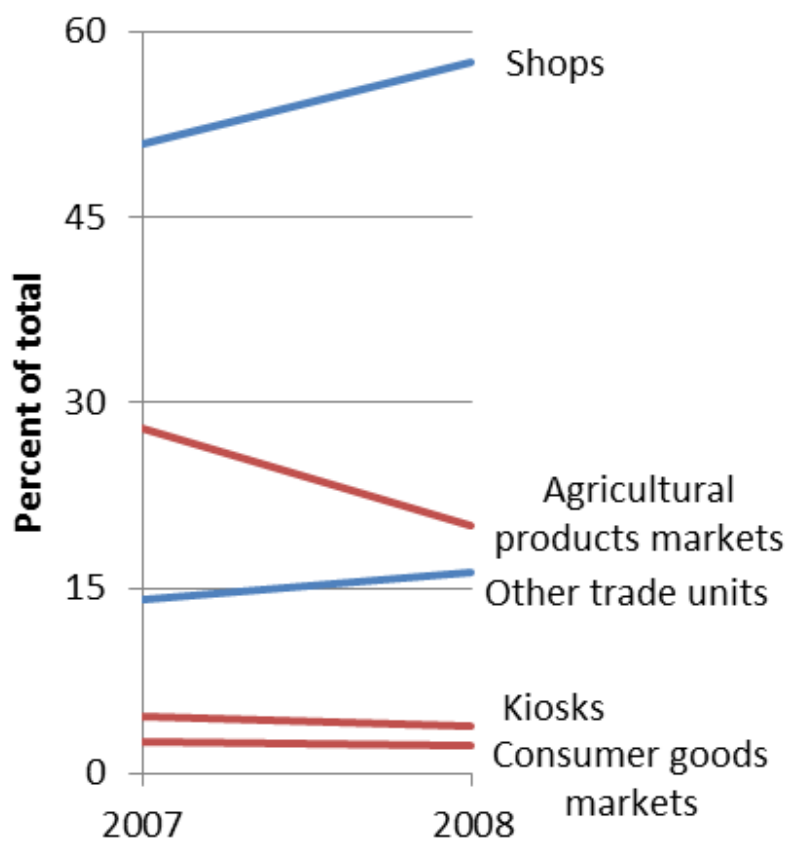
2008



- խանութների - shops - магазины
- սպառարկական ապրանքների շուկաների - agricultural products markets-рынков сельхозпродукции
- գլխավորապես սպառարկական շուկաների - consumer goods markets-рынков потребительских товаров
- կրպակների - kiosks-киосков
- արևմտյան ալ օբյեկտներ - other trade units- другие объекты торговли

(National Statistical Service of the Republic of Armenia)

## Total Volume of Retail Trade Turnover by Formation Sources

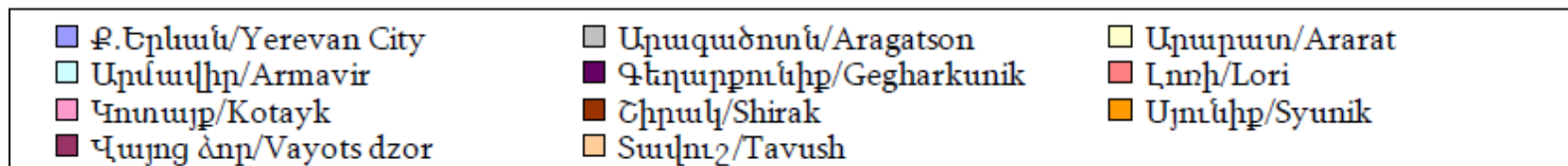
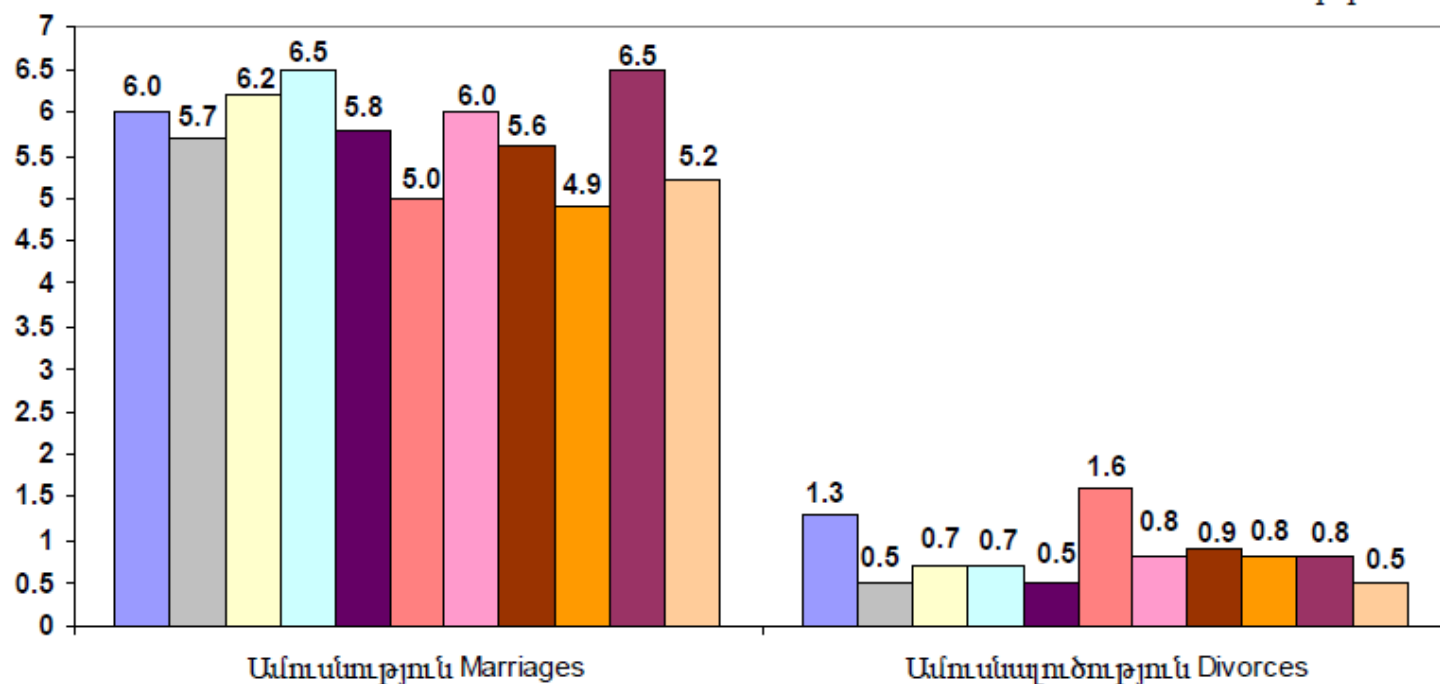


# What could be improved here?

CRUDE MARRIAGE AND DIVORCE RATE OF CITY YEREVAN AND RA MARZES, 2011

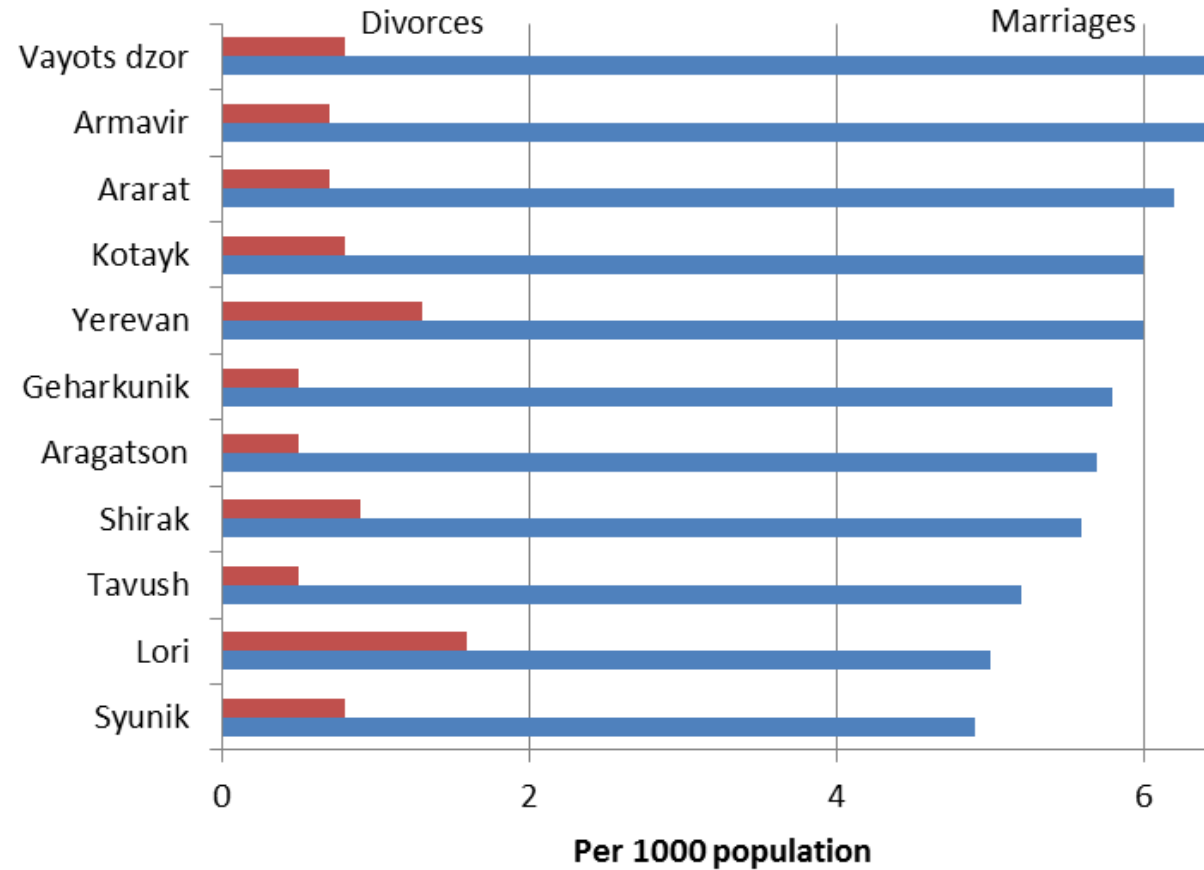
1 000 բնակչի հաշվով

Per 1 000 population



(National Statistical Service of the Republic of Armenia)

## Crude marriage and divorce rates of city Yerevan and RA marzes, 2011



# Some preattentive-processing principles

- Colors should be few, easily-distinguished, & meaningful
- Use direct labels, not legends
- Map data to an easy-to-compare perception (length > angle)
- Show comparisons directly
- Rank informatively



# Useful features of a color scheme?

# Useful features of a color scheme

- Different schemes for Qualitative vs Quantitative data
- Among Quantitative data, different schemes for Sequential vs Divergent data
- Colorblind friendly
- Print and photocopy friendly

<http://colorbrewer2.org/>

RColorBrewer package

# R demo

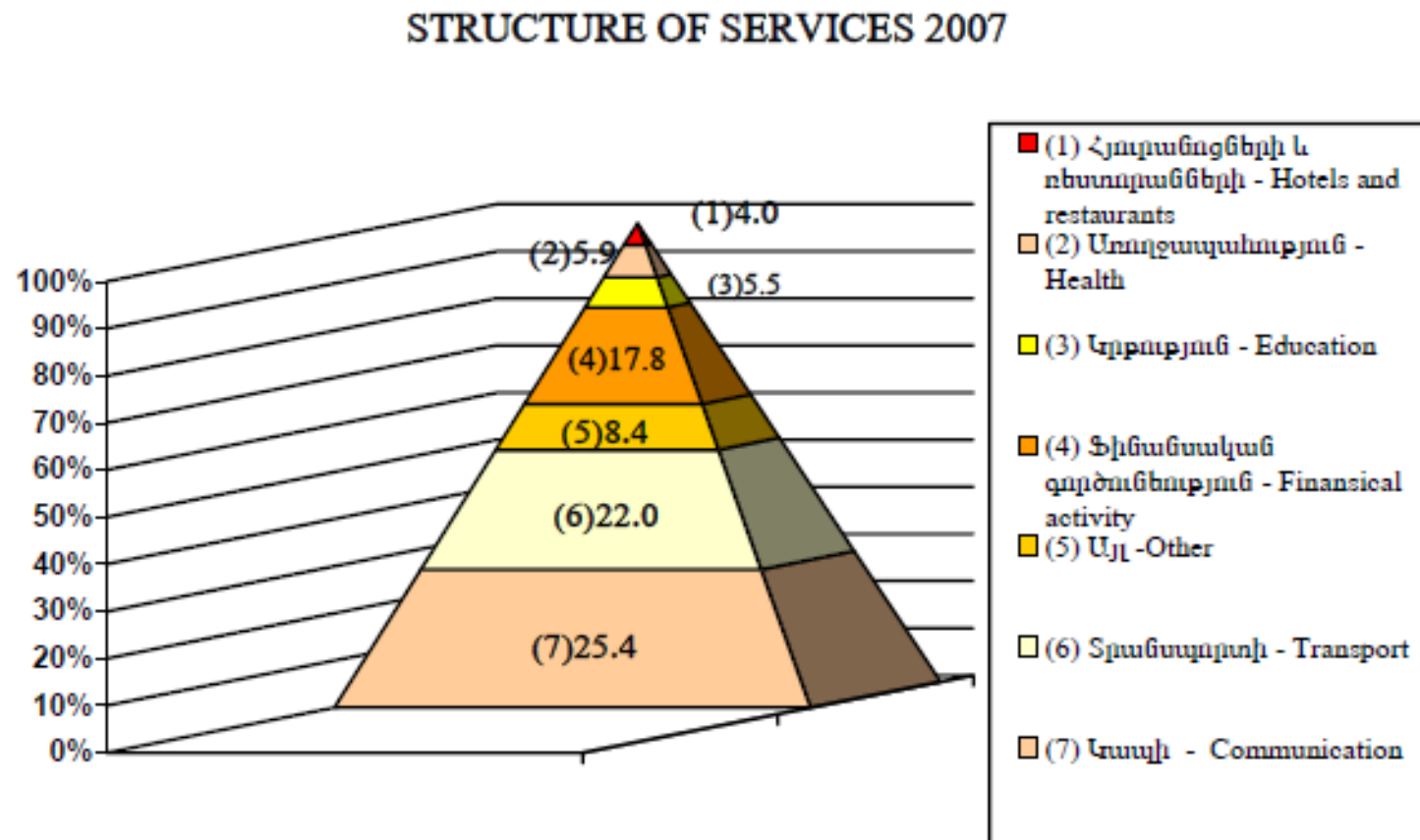
- RColorBrewer
- directlabels

<http://civilstat.com/datavis/StatBytesJune2015/>

National Health and Nutrition Examination Survey (NHANES) 2011-2012 data subset:  
infants aged 0 – 6 months, complete-cases among 3 largest race/ethnic groups.

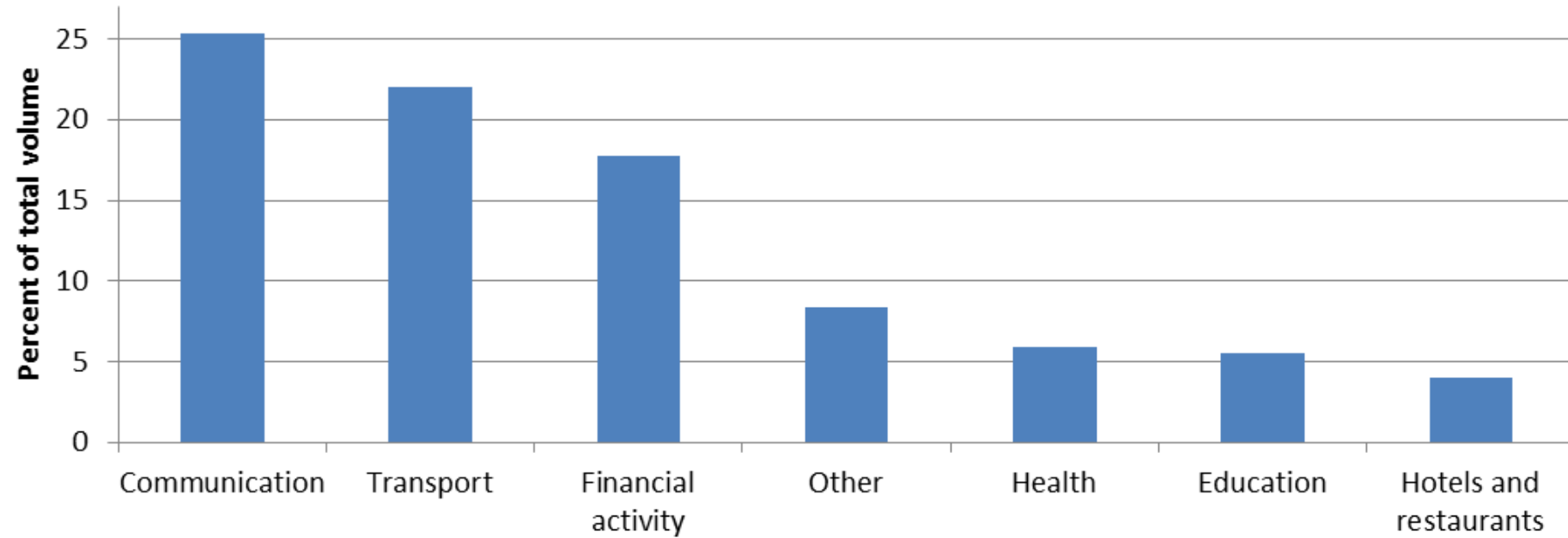
Variables: GENDER, MONTHS, RACETH, WEIGHT\_KG, LENGTH\_CM, HEAD\_CM

# What could be improved here?



(National Statistical Service of the Republic of Armenia)

## Structure of Services, 2007



# *The Grammar of Graphics, Wilkinson*

“This system is capable of producing some hideous graphics. ...  
This system cannot produce a meaningless graphic, however.”

Not a list of plot types (like Excel), but a framework for describing what to plot.

Map each data variable to a perceptual variable or **aes** (position, length, color...),  
transform it with a **stat** (identity, bin, smooth, quantile...),  
plot it as a **geom** (point, line, bar...),  
and optionally **facet** into subplots by one or two data variables

# GoG practice

Describe chart from slide 13 using Grammar of Graphics framework:

What data variables correspond to which **aes**?

What **stat** and **geom** are plotted?

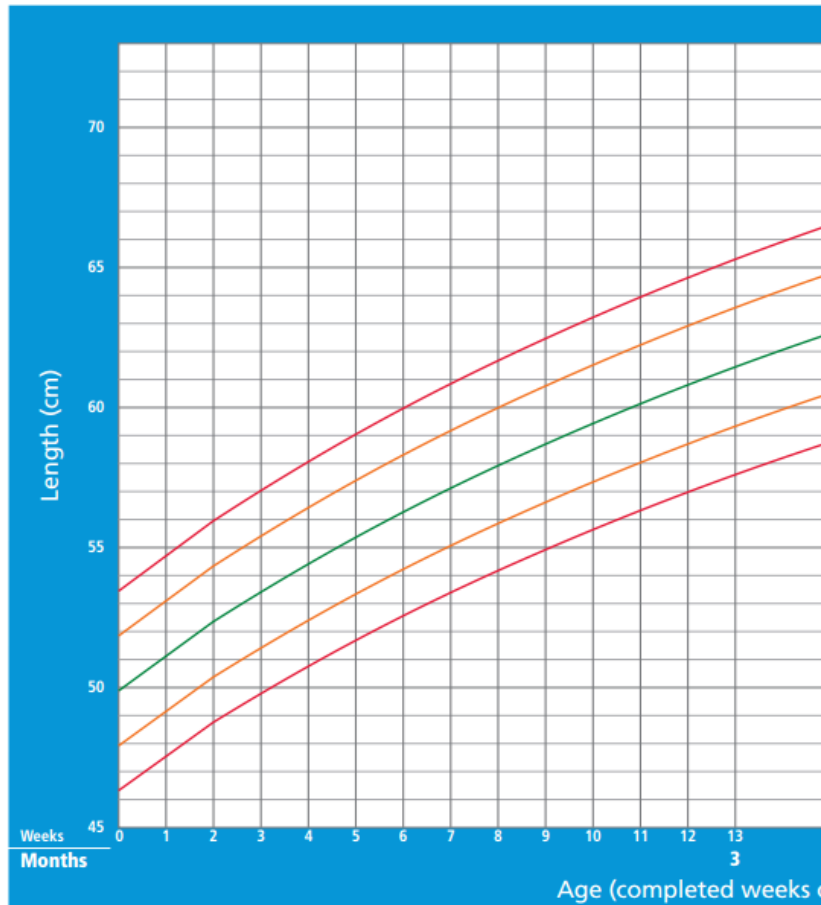
Any **facet**?

Repeat for WHO and Mayo Clinic charts on next slides

# WHO Child Growth Standards

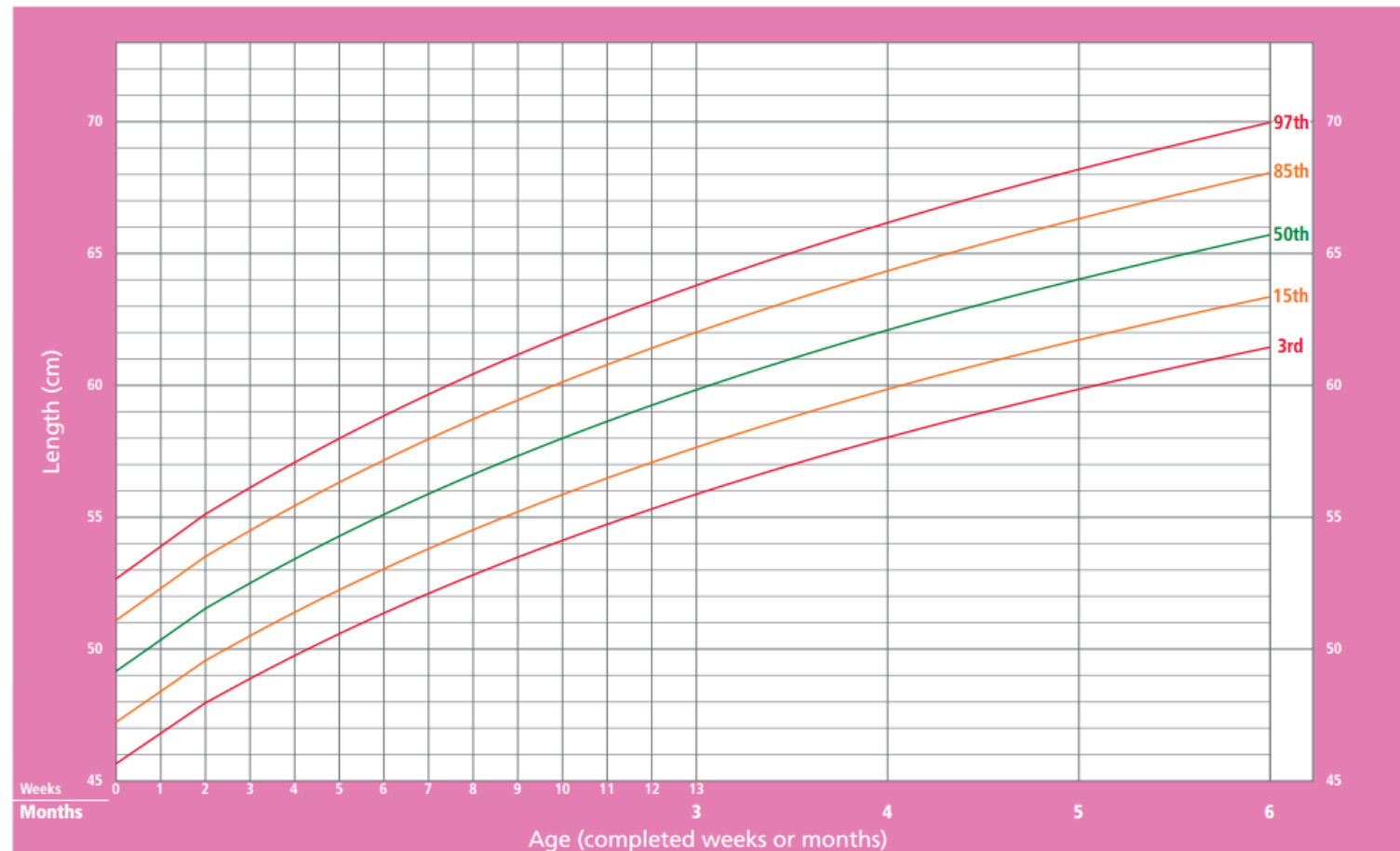
## Length-for-age BOYS

Birth to 6 months (percentiles)



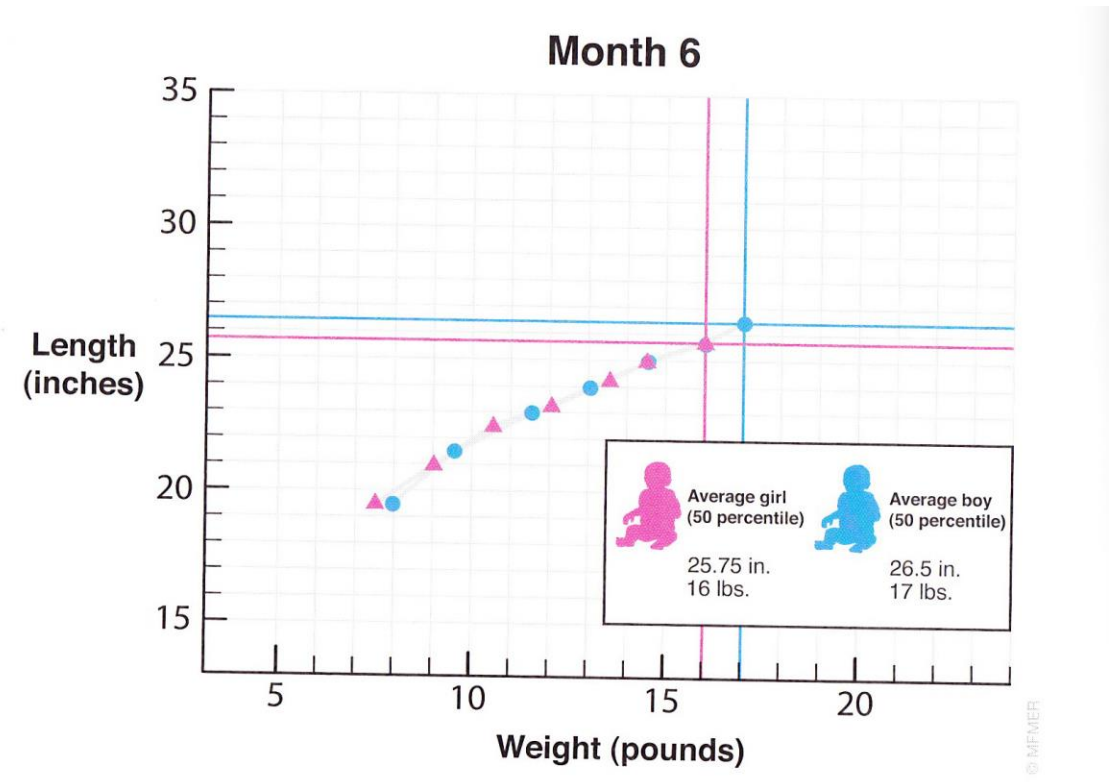
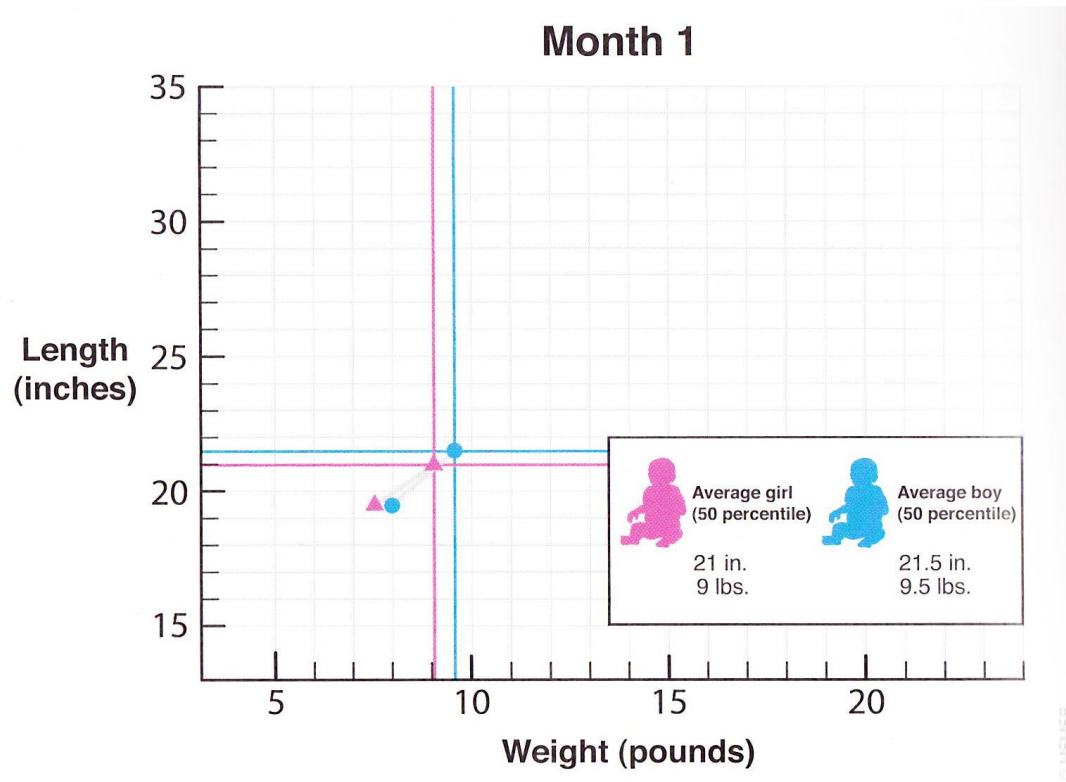
## Length-for-age GIRLS

Birth to 6 months (percentiles)





# Mayo Clinic Guide to Baby's First Year



# R demo

ggplot2

<http://docs.ggplot2.org/>

Mimic the WHO Child Growth Standards charts of  
**Length-for-age, by gender**

<http://www.who.int/childgrowth/standards/en/>

and

Mayo Clinic Guide to your Baby's First Year charts of  
**Length-for-weight, by month and gender**

# R demo

- `aes(x, y, color, linetype, size, shape, alpha)`
- `geom_point()`
- `geom_line(stat = "quantile")`
- `facet_grid(... ~ ...)`

What else can you plot with these commands and this data?

What else would you like to plot? <http://docs.ggplot2.org/>

Can you show all 6 variables at once? (Should you? 😊)