

Lecture 4: Bar Charts

February 25, 2019

Overview

Course Administration

Good, Bad and Ugly

What is a Bar Chart?

Bars and Long to Wide

Course Administration

1. Return policy brief proposals during tutorial
2. Some changes to order of events on syllabus
 - moved consultations later
 - sign up for slots April 8, 10 or 11
 - no class meeting April 15
3. Anything else?

Next Week's Good Bad and Ugly

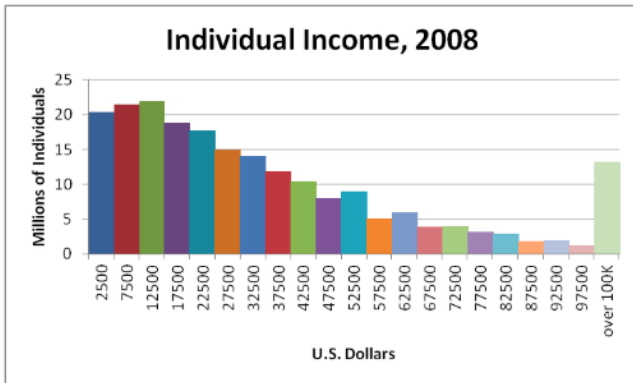
Monday by 9 am. Earlier is ok.

- ER
- WD

This Week's Good Bad and Ugly

- PH
- JB

Prusha's Example



Jessica's Example

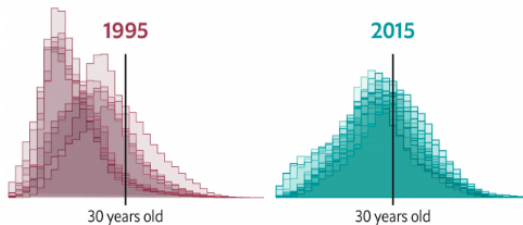
Daily chart

Europe's first-births convergence

On at least one thing, poor and rich countries in the EU now look very similar

Age of women at first birth

Selected European countries



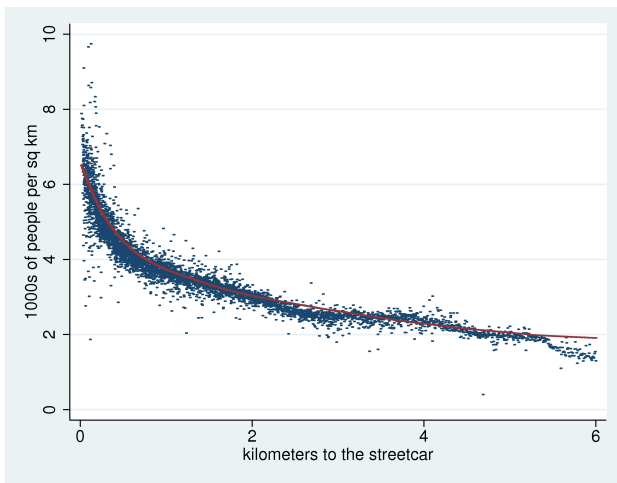
Economist.com

Few and Graph Basics

Few Reading

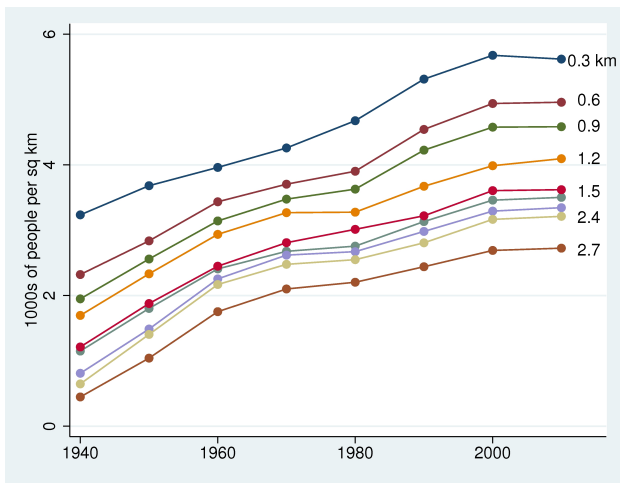
- Distributions
- Conveying information
 - position
 - hue
 - shape
- Types of information graphs can convey
- Graph selection chart
- Never-ever rules of graphs

Conveying Distributions: 2010 Structure Density by Distance to Streetcar Stop



Initially Dense Places Get Denser

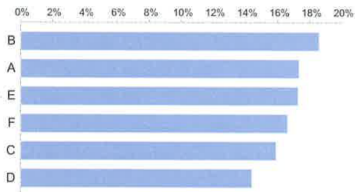
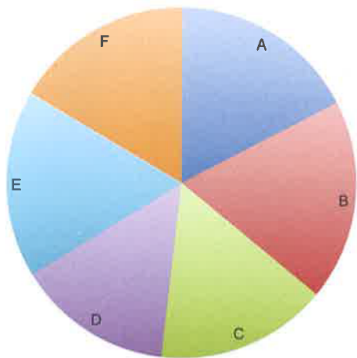
Multiple Distributions in One Graph



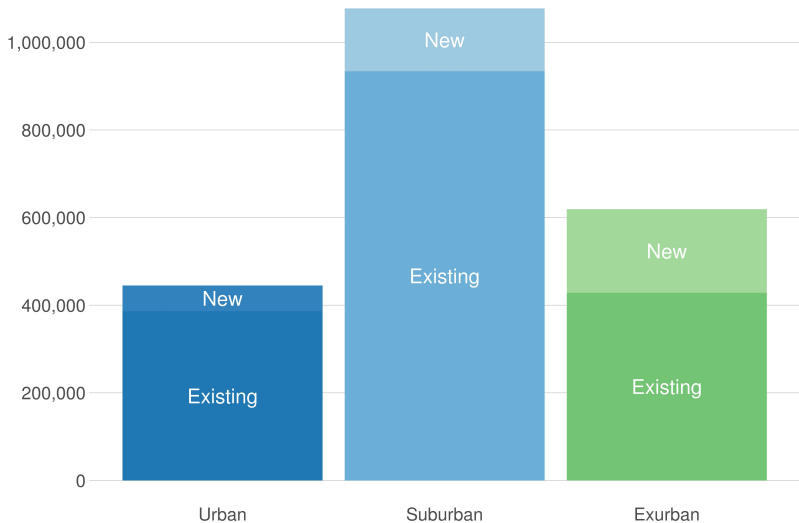
Few: Three Primary Ways to Convey Information

1. Position
2. Hue
3. Shape

When Shape Doesn't Do What You'd Hoped



Position and Hue



GRAPH CHOICE CHART

Does your question ask you...

about the **variability** of a group of data points? (i.e. the range of the data, the shape of the distribution, or what the center of the data is)

1. Do all high tides rise to the same height?
2. How variable are wind speeds in Denmark?
3. What is the range and distribution of incomes in Sudan?

to **compare two or more groups** to decide if the groups are the same or different?

if **two numeric factors are correlated**

1. Is the temperature inside the house correlated with the temperature outside?
2. How did electricity used by the kitchen circuit fluctuate during the past week?

how a **total is proportioned** into sub-groups? (Or what proportion a sub-group is of a total?)

1. What were Brazil's most significant exports in 2015?
2. What proportion of global electricity production comes from wind?
3. How do Parisians typically commute to work?

Do you want to compare the **variability of all data points** in each group to decide if any difference between the groups is meaningful?

1. Which of the two solar cars consistently goes the farthest?
2. Is there a meaningful difference in the heights of fertilized and unfertilized bean plants?

Are you comparing **single numbers** that summarize a group? (such as mean, median, or total...)

1. Was the total snowfall greater this winter than last winter?
2. Do cats and dogs have the same average body temperature?
3. How do the median incomes for the US and India compare?

Does it ask about how something changes through **linear TIME**?

N
Y

1. Is the fuel efficiency of a car related to its weight?
2. Are smoking rates correlated with median income?
3. Given a fixed volume, how are temperature and pressure related?

1. Is sea level rising?
2. How did my weight change over the last 3 months?

Frequency Plot

FOR EACH GROUP MAKE A

MAKE EITHER

MAKE EITHER

Histogram



Box Plots



Dot Plot



Bar Graph



Scatter Plot



Line Graph



Pie Chart



Stacked Bar Chart



Two Fundamental Rules of Graphs

1. No 3D
2. Never start bars not at zero
3. Sparingly start line graphs above zero

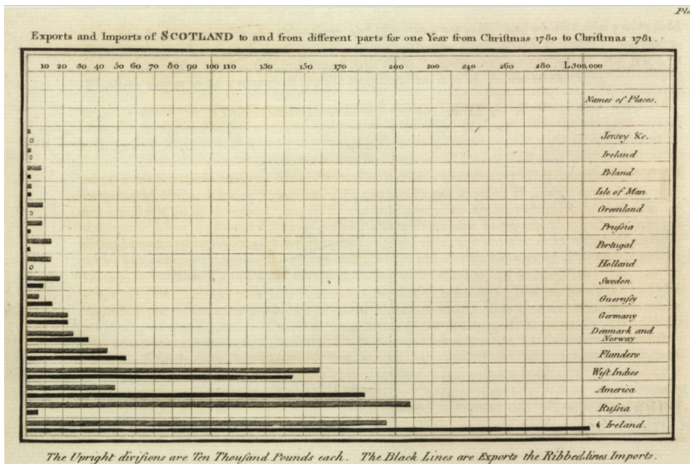
Why the difference between 2 and 3?

Bar Charts

Bar Chart

- Compare levels across categories
- Compare levels of shares across categories
- Emphasize rank order of levels
- Highlight one level relative to others

The First Bar Chart

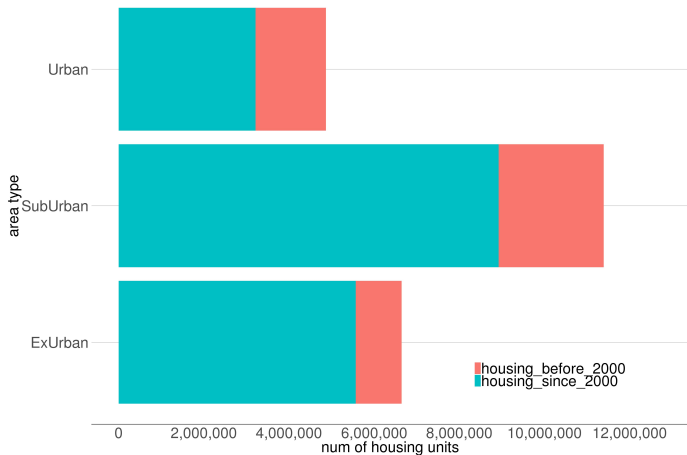


Playfair, William, 1786. *The Commercial and Political Atlas*

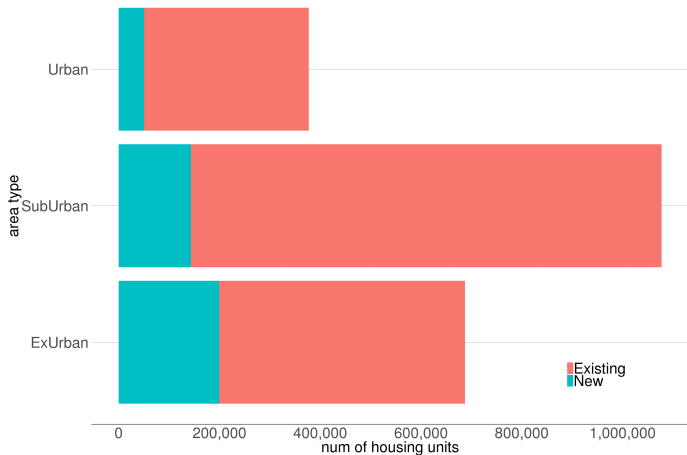
General Principles for Bar Charts

- Use horizontal bars when you have long labels
- Rank when you want to highlight ordering
- Avoid color as decoration
- As much as possible, put legend directly on the graph

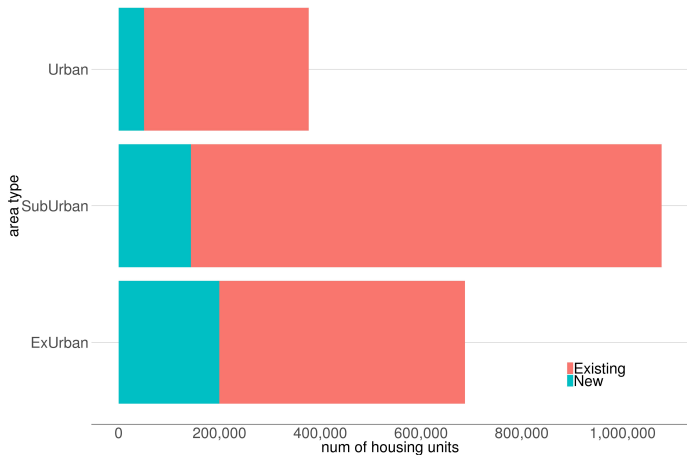
Going From a Bad Bar Chart to a Decent One



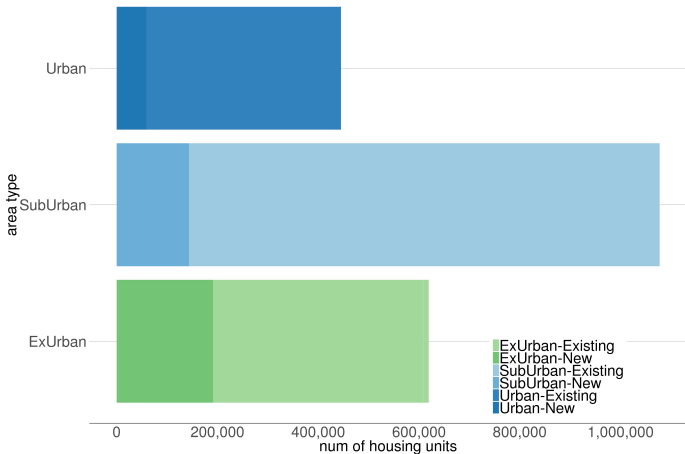
Going From a Bad Bar Chart to a Decent One



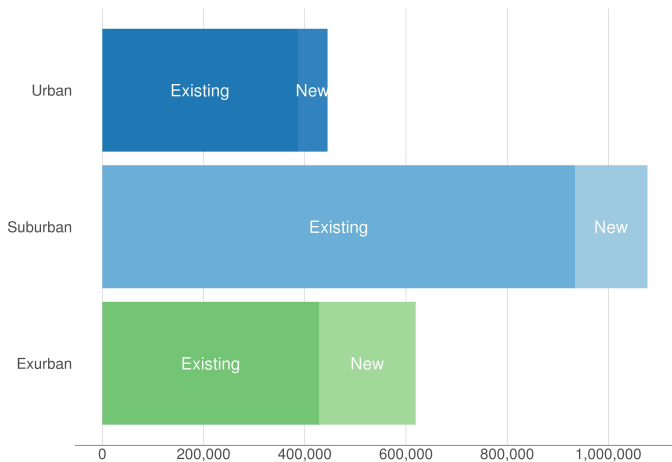
Going From a Bad Bar Chart to a Decent One



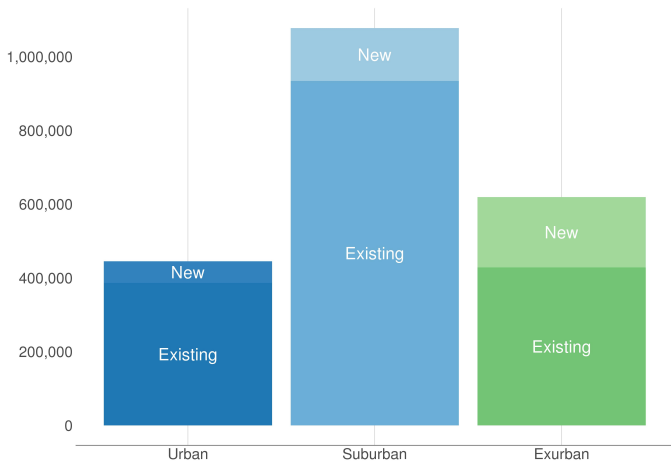
Going From a Bad Bar Chart to a Decent One



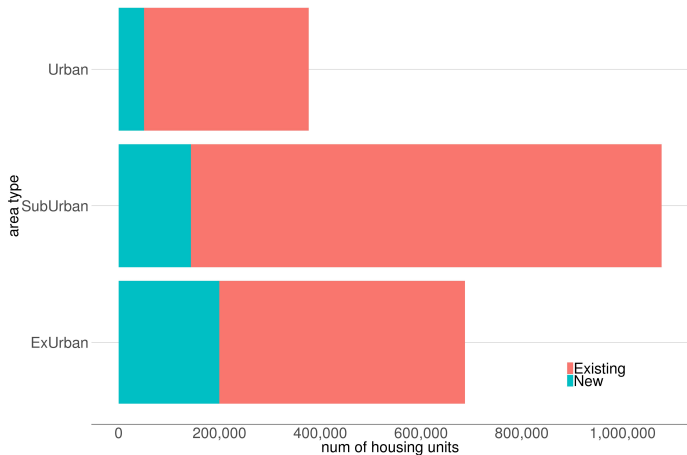
Going From a Bad Bar Chart to a Decent One



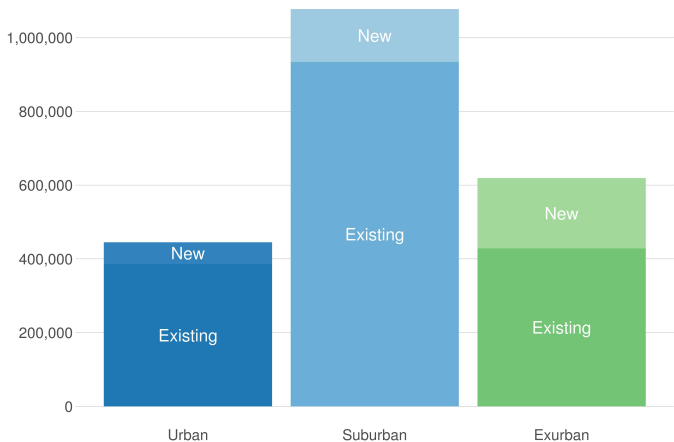
Going From a Bad Bar Chart to a Decent One



Going From a Bad Bar Chart to a Decent One



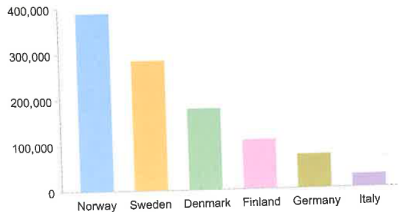
Going From a Bad Bar Chart to a Decent One



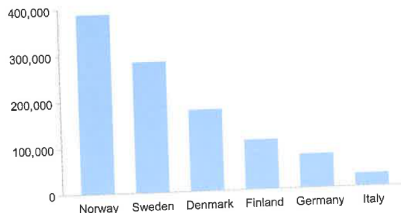
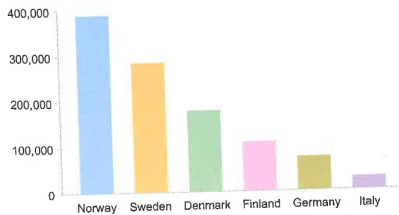
Two Examples

- Bars with colors
- Ranked bars

Bars with Colors

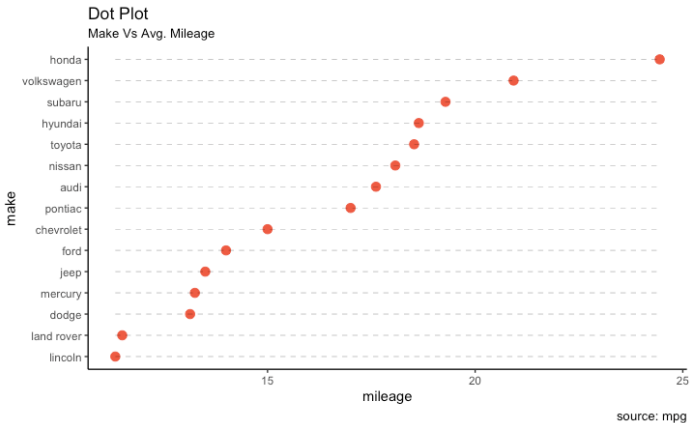


Bars with Colors



Few Figure 10.22

Ranked Bars



See source for this [here](#).

R

Today

- A. Bar chart syntax
- B. Wide vs long data
- C. Factor levels
- D. `dplyr`'s `mutate` command

A.1. Bar Chart Syntax

At its most basic

```
new.graph <- ggplot() +  
  geom_bar(data = [your data],  
           mapping = aes(x = [categorical variable],  
                         y = [value]))
```

This should look familiar from last week

A.2. Bar Chart Additions

Of course, there are many more things you can do

- ▶ make R use the value in the y variable: `geom_col()`
- ▶ make stacked bars: `position = "stack"`
- ▶ make grouped bars: `position = "dodge"`
- ▶ change the bar width
- ▶ change bar colors
- ▶ put labels on bars
- ▶ and still oodles more

B.1 Wide vs long data

Wide data

```
wide
```

```
##   state temp_2018 temp_2017
## 1    VA         108         100
## 2    MD         103         203
## 3    DC         102         105
```

ggplot prefers long data. With these wide data you can make two charts, but you can't put both years on one chart.

B.2. Long data

```
long[,c("state", "year", "temp")]
```

```
##   state year temp
## 1    VA 2017  100
## 2    MD 2017  203
## 3    DC 2017  105
## 4    VA 2018  108
## 5    MD 2018  103
## 6    DC 2018  102
```

B.3. How to get from one to another

Use the `gather()` command from `dplyr`

```
library(tidyr)
long <- gather(data = wide,
               key = org.var,
               value = temp,
               c("temp_2017", "temp_2018"))
long$year <- substr(long$org.var, start = 6, stop = 9)
long
```

B.4. Final product

```
long
```

```
##   state  org.var temp year
## 1    VA temp_2017  100 2017
## 2    MD temp_2017  203 2017
## 3    DC temp_2017  105 2017
## 4    VA temp_2018  108 2018
## 5    MD temp_2018  103 2018
## 6    DC temp_2018  102 2018
```


C.1. Factor levels

- ▶ we particularly care about factor levels this class
- ▶ R orders bar charts by the order of the factor
- ▶ to change the order, change the order of the factor

C.2. Setting up a factor variable

```
str(long)
```

```
## 'data.frame':    6 obs. of  4 variables:
## $ state  : Factor w/ 3 levels "DC","MD","VA": 3 2 1 3 2 1
## $ org.var: chr  "temp_2017" "temp_2017" "temp_2017" "temp_2017"
## $ temp   : num  100 203 105 108 103 102
## $ year   : chr  "2017" "2017" "2017" "2018" ...
```

- ▶ state is a factor variable
- ▶ has three levels: DC, MD, VA
- ▶ in that order – R auto-alphabetizes
- ▶ suppose we prefer it in another order: VA, DC, MD

C.3. Re-ordering a factor

```
levels(long$state)
```

```
## [1] "DC" "MD" "VA"
```

```
long$state <- factor(long$state,  
                     levels = c("VA", "DC", "MD"))  
levels(long$state)
```

```
## [1] "VA" "DC" "MD"
```

D.0 Call dplyr package

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

D.1. mutate()

- ▶ if you know stata's egen, it's like that
- ▶ in wide dataframe above, add a row that is total temperature

```
library(dplyr)
```

```
wide2 <- mutate(.data = wide,  
                tot.temp.2017=sum(temp_2017))
```

```
wide2
```

```
##   state temp_2018 temp_2017 tot.temp.2017  
## 1    VA         108         100          408  
## 2    MD         103         203          408  
## 3    DC         102         105          408
```

- ▶ it does many many other things as well
- ▶ you can use all kinds of functions in the second term