

Problem with Data Visualization

lecture3

We have a problem

We have a problem

Why does this happen?

SOURCE: THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION, E. TUFTS, 1983.

1

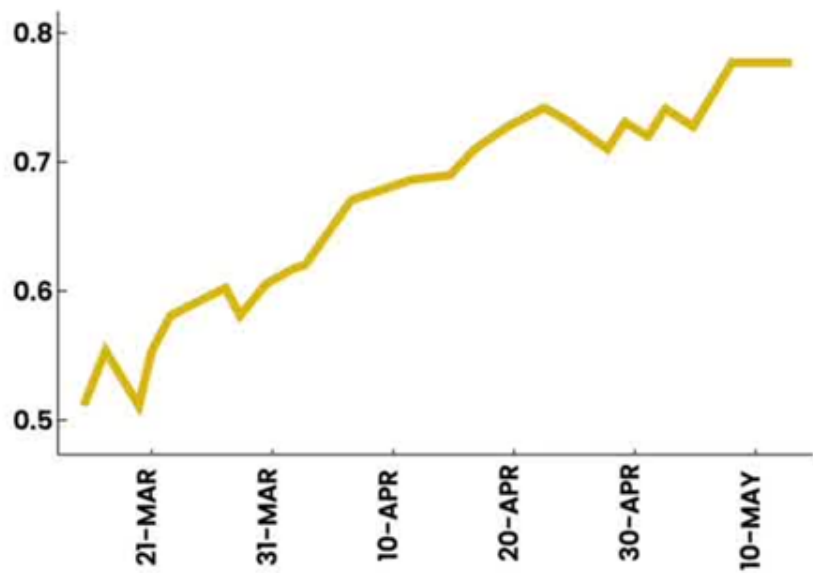
The (historic) lack of quantitative skills of professional artists

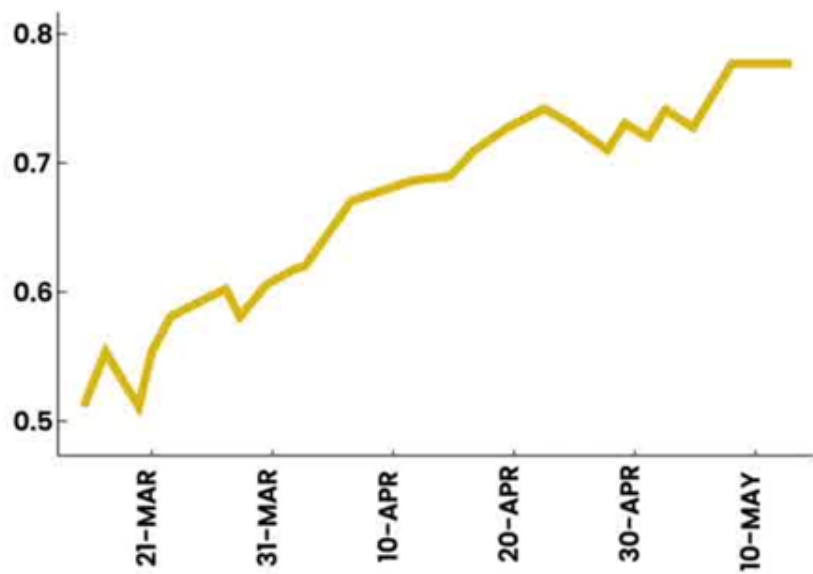
- Most of the people who produce graphs are artists, designers, developers, that have no training in the analysis of data.
- Often, their goal is to beautify data, even at the expense of statistical integrity.

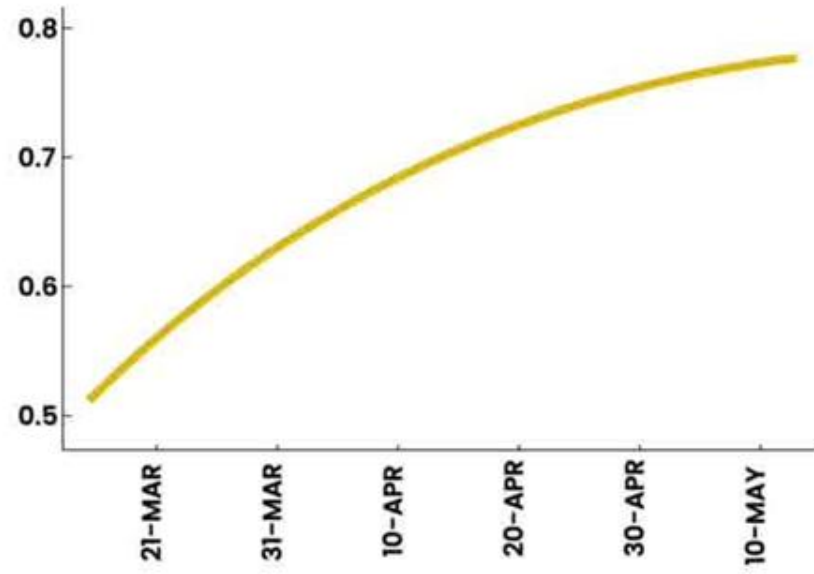
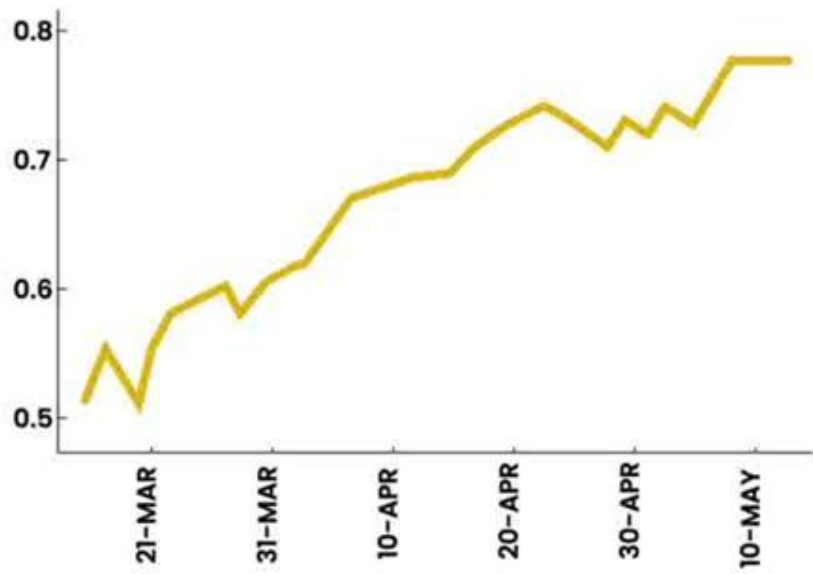
1

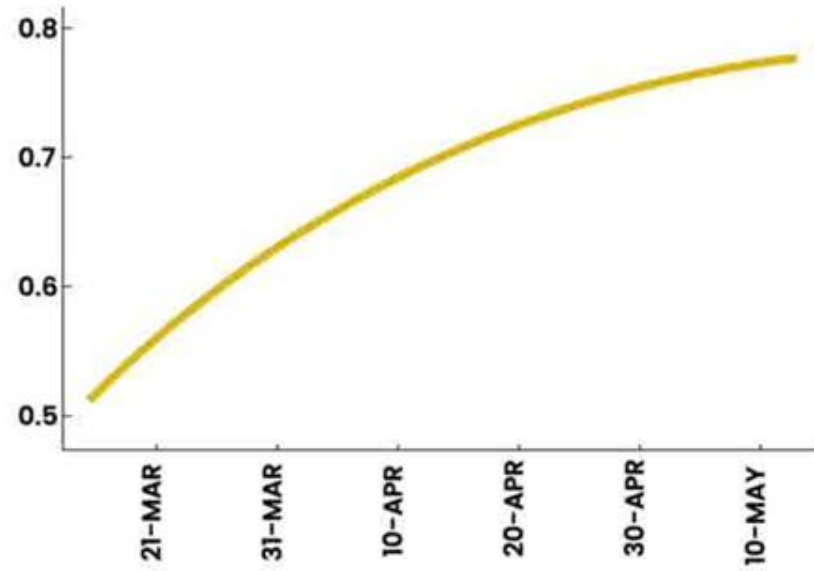
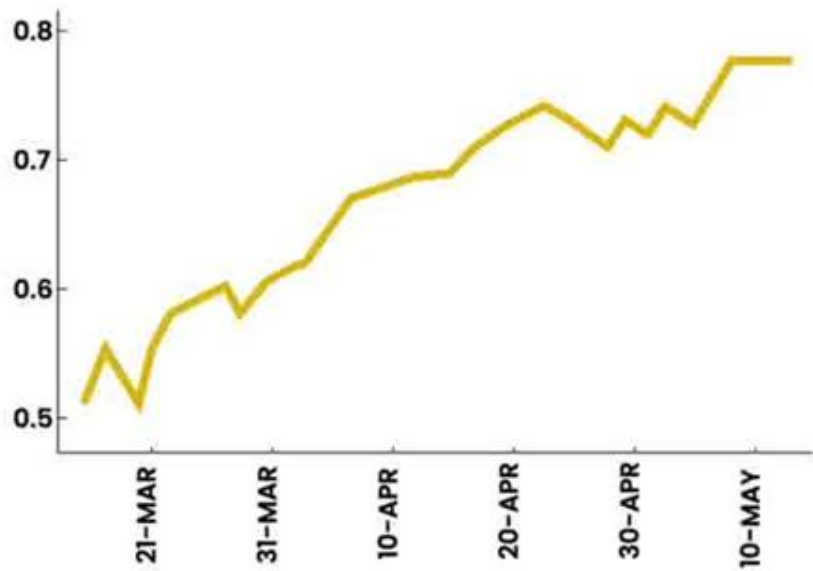
The (historic) lack of quantitative skills of professional artists

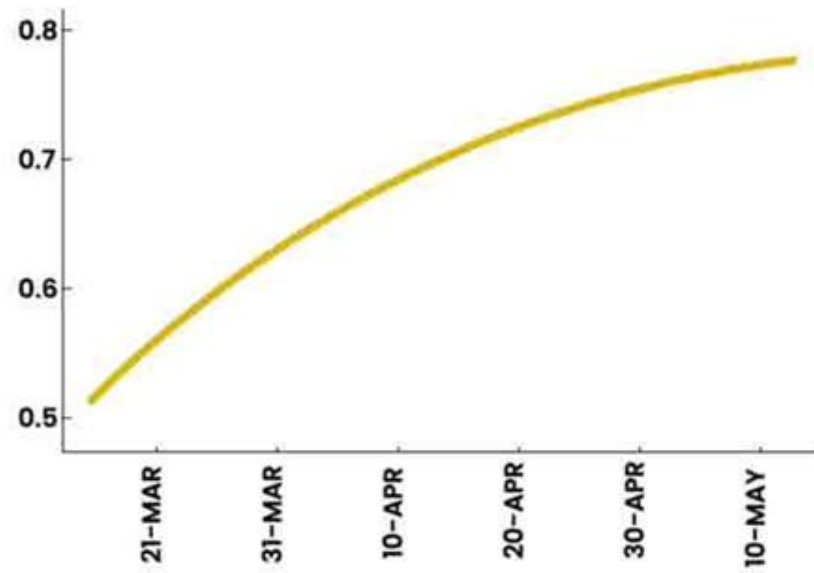
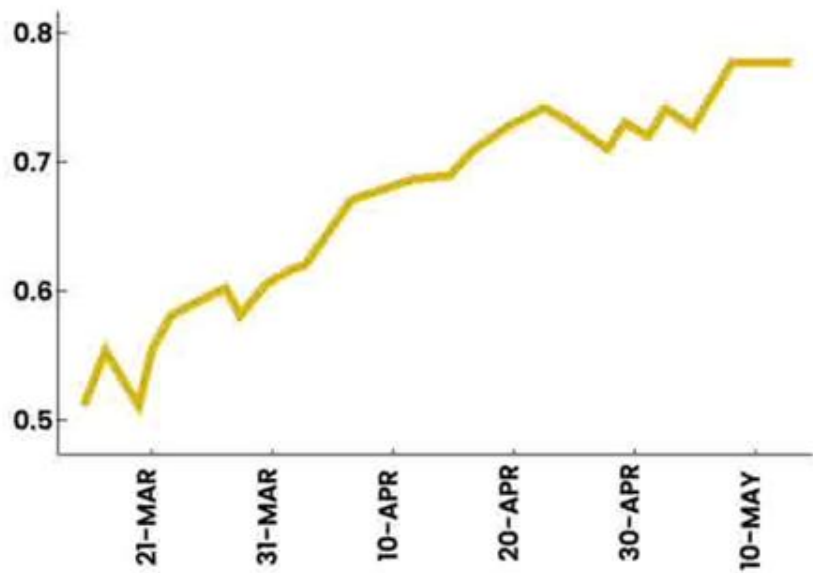
- Most of the people who produce graphs are artists, designers, developers, that have no training in the analysis of data.
- Often, their goal is to beautify data, even at the expense of statistical integrity.











1**The (historic) lack of quantitative skills of professional artists**

- Most of the people who produce graphs are artists, designers, developers, that have no training in the analysis of data.
- Often, their goal is to beautify data, even at the expense of statistical integrity.

2**The belief that statistical data is boring to the user**

1**The (historic) lack of quantitative skills of professional artists**

- Most of the people who produce graphs are artists, designers, developers, that have no training in the analysis of data.
- Often, their goal is to beautify data, even at the expense of statistical integrity.

2**The belief that statistical data is boring to the user**

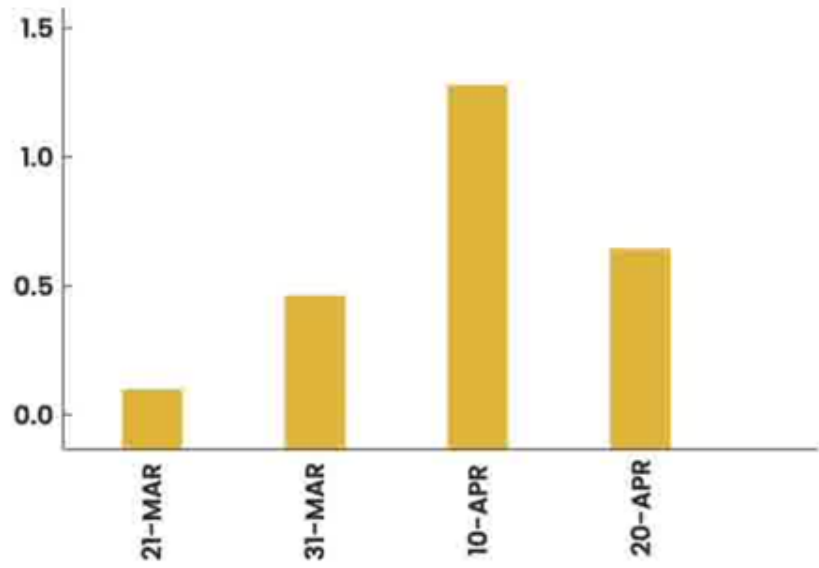
- There is a common belief that data is boring.
- This produces graphics with embellishments that distract from the data. If your data is boring, then you've got the wrong numbers.

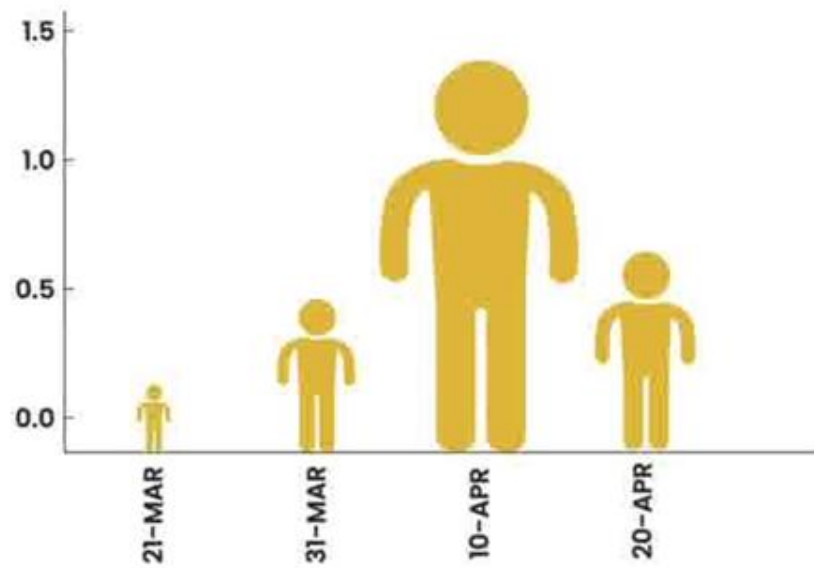
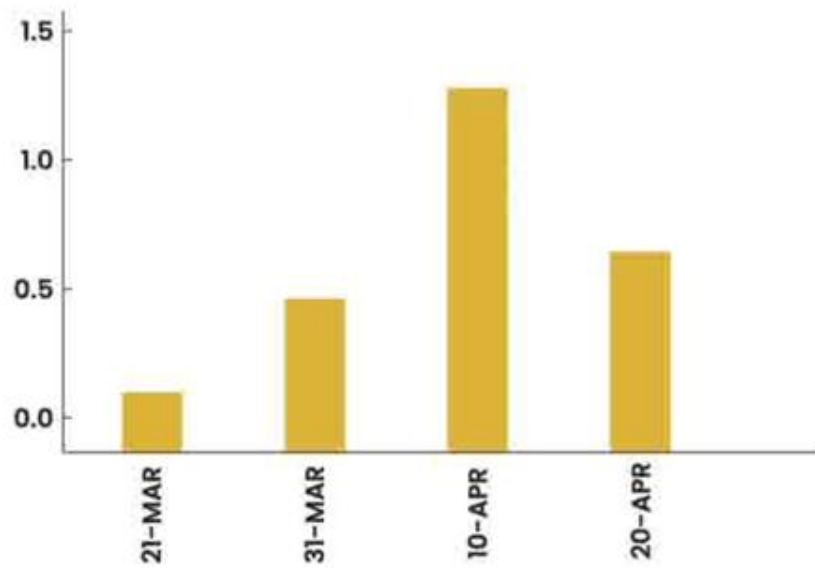
1**The (historic) lack of quantitative skills of professional artists**

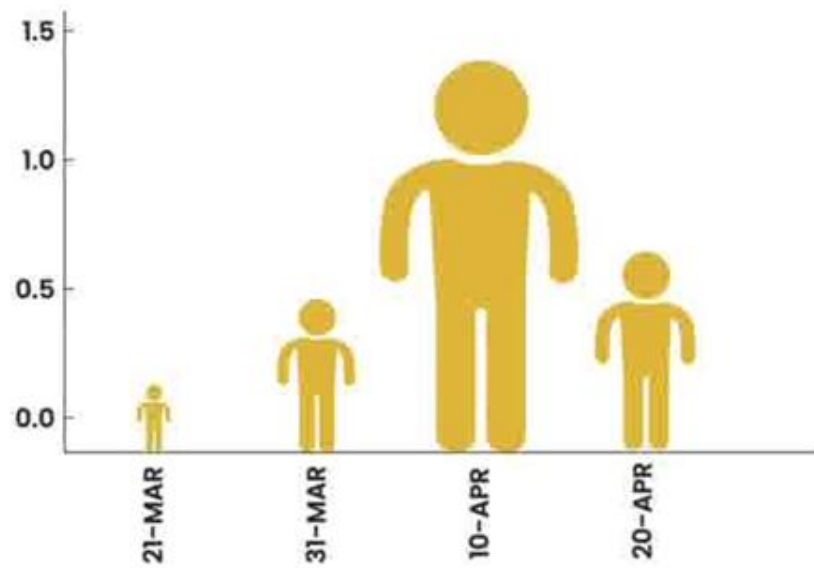
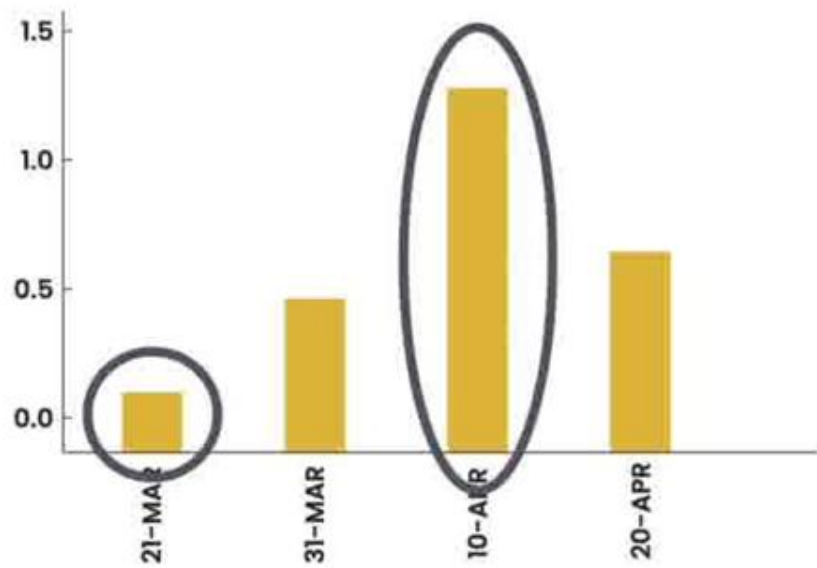
- Most of the people who produce graphs are artists, designers, developers, that have no training in the analysis of data.
- Often, their goal is to beautify data, even at the expense of statistical integrity.

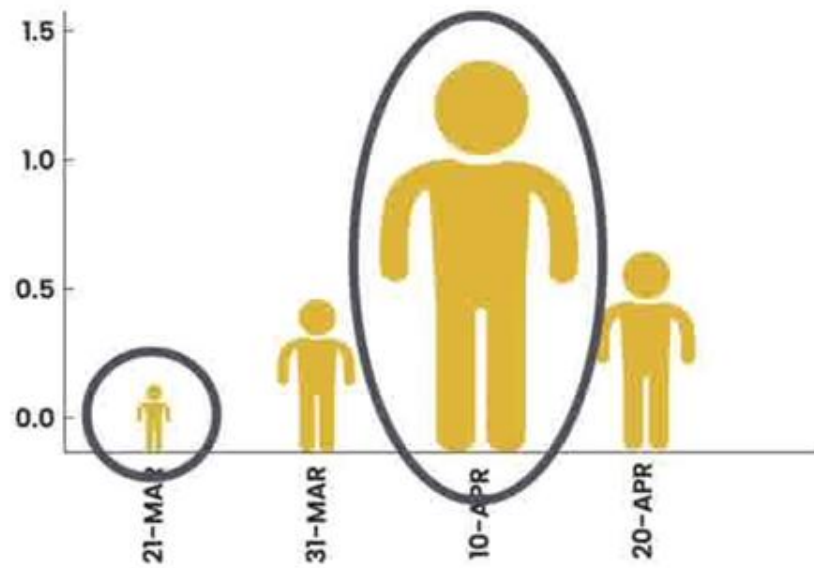
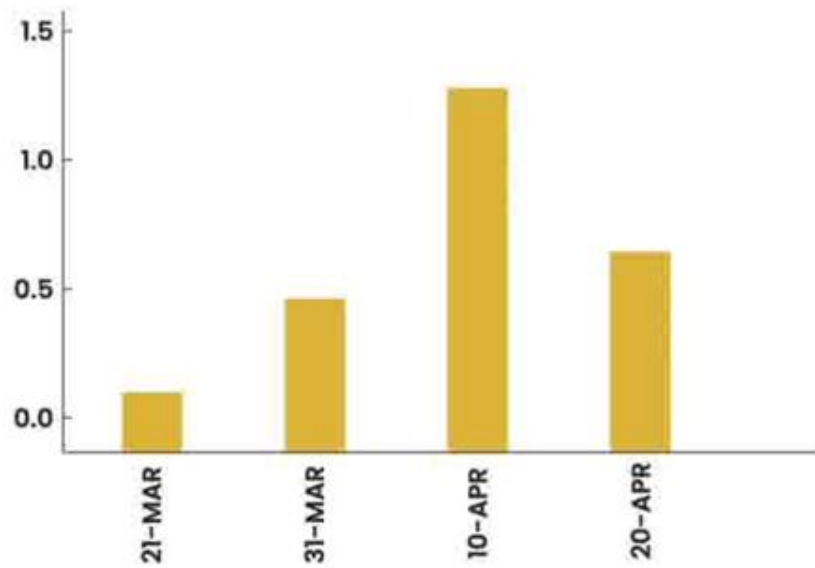
2**The belief that statistical data is boring to the user**

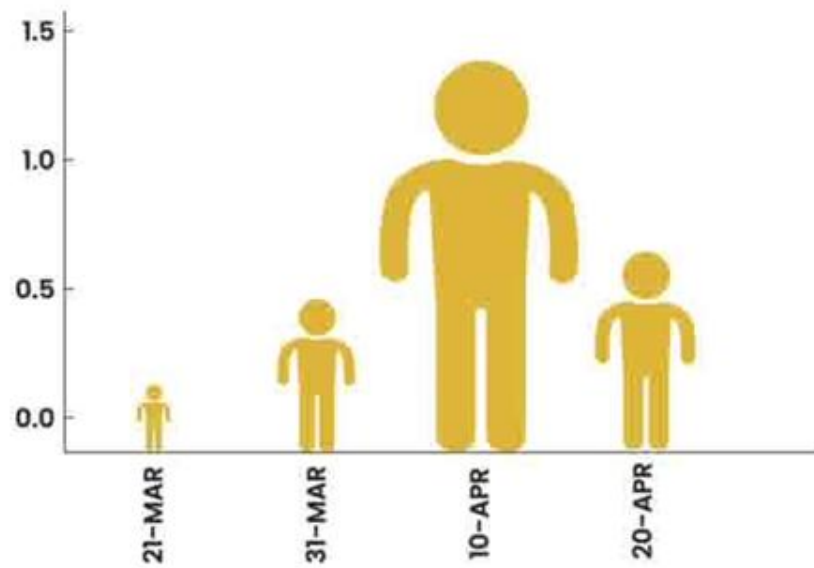
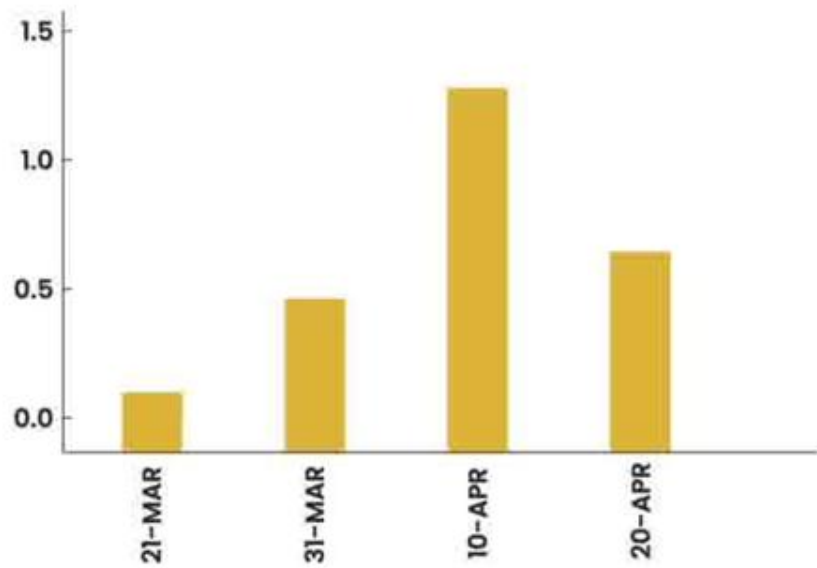
- There is a common belief that data is boring.
- This produces graphics with embellishments that distract from the data. If your data is boring, then you've got the wrong numbers.











1 The (historic) lack of quantitative skills of professional artists

- Most of the people who produce graphs are artists, designers, developers, that have no training in the analysis of data.
- Often, their goal is to beautify data, even at the expense of statistical integrity.

2 The belief that statistical data is boring to the user

- There is a common belief that data is boring.
- This produces graphics with embellishments that distract from the data. If your data is boring, then you've got the wrong numbers.

3 The belief that the audience won't understand charts

1 The (historic) lack of quantitative skills of professional artists

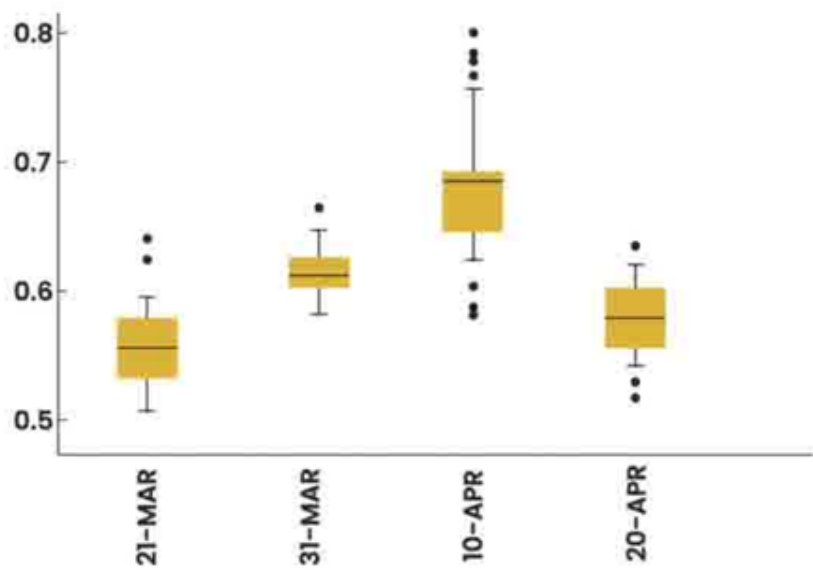
- Most of the people who produce graphs are artists, designers, developers, that have no training in the analysis of data.
- Often, their goal is to beautify data, even at the expense of statistical integrity.

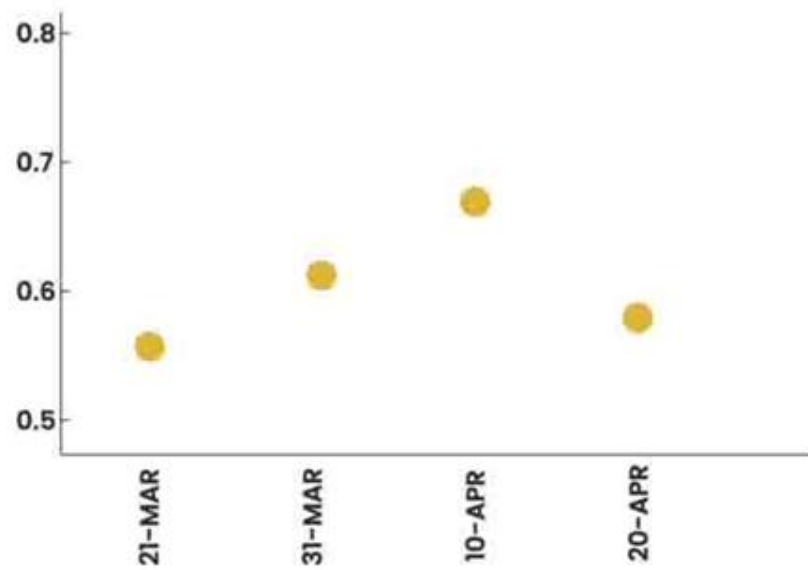
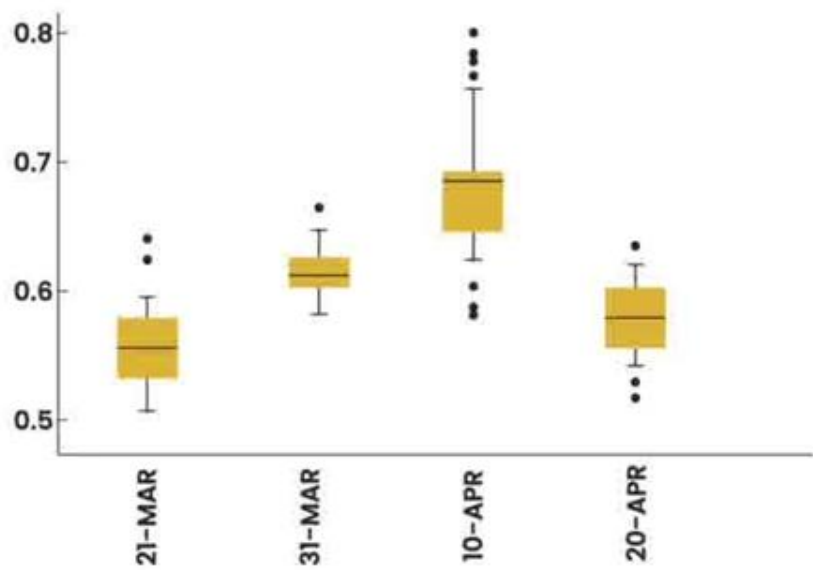
2 The belief that statistical data is boring to the user

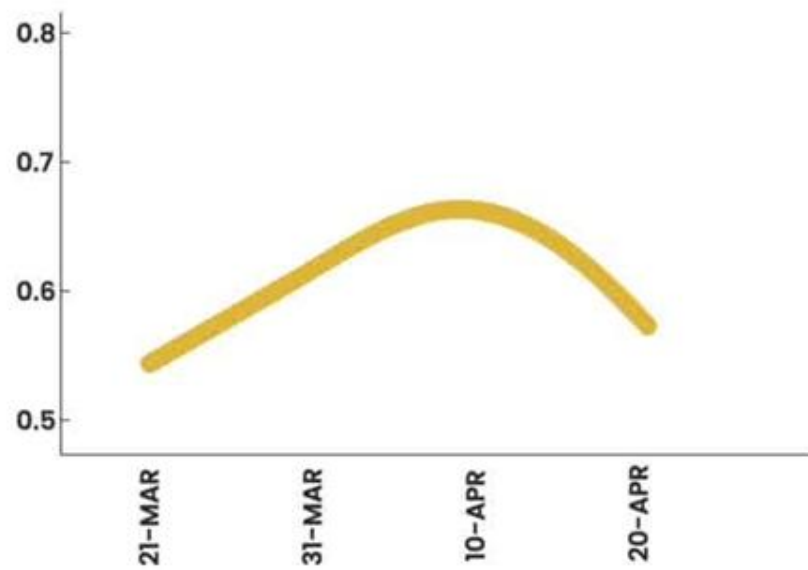
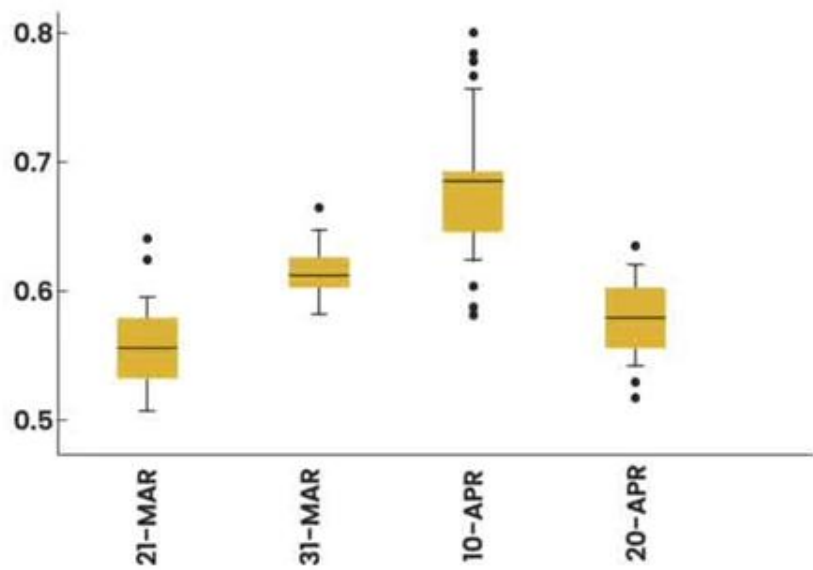
- There is a common belief that data is boring.
- This produces graphics with embellishments that distract from the data. If your data is boring, then you've got the wrong numbers.

3 The belief that the audience won't understand charts

- It is true that getting used to reading graphs takes some practice, but your goal is to help your audience read yours, by making it crystal clear, not by simplifying its contents until you've got an unnecessary graph.







1 The (historic) lack of quantitative skills of professional artists

- Most of the people who produce graphs are artists, designers, developers, that have no training in the analysis of data.
- Often, their goal is to beautify data, even at the expense of statistical integrity.

2 The belief that statistical data is boring to the user

- There is a common belief that data is boring.
- This produces graphics with embellishments that distract from the data. If your data is boring, then you've got the wrong numbers.

3 The belief that the audience won't understand charts

- It is true that getting used to reading graphs takes some practice, but your goal is to help your audience read yours, by making it crystal clear, not by simplifying its contents until you've got an unnecessary graph.

ORIGINAL DATA

ORIGINAL DATA



SIMPLIFICATION OF THE DATA

ORIGINAL DATA



SIMPLIFICATION OF THE DATA



IRRELEVANT DATA

ORIGINAL DATA



SIMPLIFICATION OF THE DATA



IRRELEVANT DATA



IRRELEVANT PLOT



ATTEMPTS TO BEAUTIFY

ORIGINAL DATA



SIMPLIFICATION OF THE DATA



IRRELEVANT DATA



IRRELEVANT PLOT



ATTEMPTS TO BEAUTIFY



CHARTJUNK PLOT

ORIGINAL DATA



SIMPLIFICATION OF THE DATA



IRRELEVANT DATA



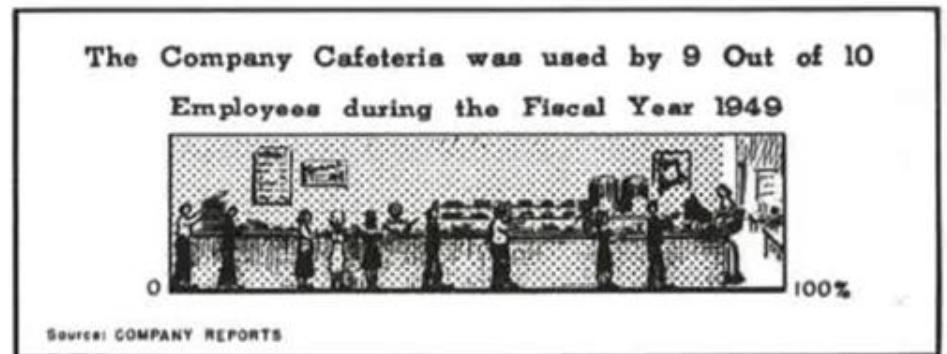
IRRELEVANT PLOT



ATTEMPTS TO BEAUTIFY



CHARTJUNK PLOT



SOURCE: MARY ELEANOR SPEAR, CHARTING STATISTICS, EXTRACTED FROM THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION, E.TUFTE.

A big part of the problem

A big part of the problem

Plotting is done in autopilot mode

A big part of the problem

Plotting is done in autopilot mode

Technological achievements make us forget that drawing a plot is a rational, creative process, nothing automatic.

A big part of the problem

Plotting is done in autopilot mode

Technological achievements make us forget that drawing a plot is a rational, creative process, nothing automatic.

We don't plan our plots ahead, instead we expect technology to give us the solution

A big part of the problem

The screenshot shows a Microsoft Excel spreadsheet titled "Financial Sample". The ribbon includes Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, and Table. The active cell is K22, containing the formula =1101. The spreadsheet displays a table with the following data:

Segment	Country	Product	Discount Base	Units Sold	Manufacturing	Sale Price	Sales Value	Discounts
Government	Canada	Carreters	None	1618.5	\$ 3.00	\$ 20.00	\$ 32370.00	\$ -
Government	Germany	Carreters	None	1321	\$ 8.00	\$ 20.00	\$ 26420.00	\$ -
Midmarket	France	Carreters	None	2178	\$ 3.00	\$ 15.00	\$ 32670.00	\$ -
Midmarket	Germany	Carreters	None	888	\$ 3.00	\$ 15.00	\$ 13320.00	\$ -
Midmarket	Mexico	Carreters	None	2470	\$ 3.00	\$ 15.00	\$ 37050.00	\$ -
Government	Germany	Carreters	None	1513	\$ 3.00	\$ 350.00	\$ 529550.00	\$ -
Midmarket	Germany	Montana	None	921	\$ 5.00	\$ 15.00	\$ 13815.00	\$ -
Channel Partners	Canada	Montana	None	2518	\$ 5.00	\$ 12.00	\$ 30216.00	\$ -
Government	France	Montana	None	1899	\$ 5.00	\$ 20.00	\$ 37980.00	\$ -
Channel Partners	Germany	Montana	None	1545	\$ 5.00	\$ 12.00	\$ 18540.00	\$ -
Midmarket	Mexico	Montana	None	2470	\$ 5.00	\$ 13.00	\$ 37050.00	\$ -
Enterprise	Canada	Montana	None	2665.5	\$ 5.00	\$ 125.00	\$ 333187.50	\$ -
Small Business	Mexico	Montana	None	958	\$ 5.00	\$ 300.00	\$ 287400.00	\$ -
Government	Germany	Montana	None	2148	\$ 5.00	\$ 7.00	\$ 15022.00	\$ -
Enterprise	Canada	Montana	None	345	\$ 5.00	\$ 125.00	\$ 43125.00	\$ -
Midmarket	United States of America	Montana	None	615	\$ 5.00	\$ 15.00	\$ 9225.00	\$ -
Government	Canada	Paseo	None	292	\$ 10.00	\$ 20.00	\$ 5840.00	\$ -
Midmarket	Mexico	Paseo	None	974	\$ 10.00	\$ 15.00	\$ 14610.00	\$ -
Channel Partners	Canada	Paseo	None	2518	\$ 10.00	\$ 12.00	\$ 30216.00	\$ -
Government	Germany	Paseo	None	1006	\$ 10.00	\$ 350.00	\$ 352100.00	\$ -
Channel Partners	Germany	Paseo	None	367	\$ 10.00	\$ 12.00	\$ 4404.00	\$ -
Government	Mexico	Paseo	None	883	\$ 10.00	\$ 7.00	\$ 6181.00	\$ -
Midmarket	France	Paseo	None	548	\$ 10.00	\$ 15.00	\$ 8235.00	\$ -
Small Business	Mexico	Paseo	None	788	\$ 10.00	\$ 300.00	\$ 236400.00	\$ -
Midmarket	Mexico	Paseo	None	2472	\$ 10.00	\$ 15.00	\$ 37080.00	\$ -
Government	United States of America	Paseo	None	1143	\$ 10.00	\$ 7.00	\$ 8001.00	\$ -
Government	Canada	Paseo	None	1725	\$ 10.00	\$ 350.00	\$ 603750.00	\$ -
Channel Partners	United States of America	Paseo	None	912	\$ 10.00	\$ 12.00	\$ 10944.00	\$ -
Midmarket	Canada	Paseo	None	2152	\$ 10.00	\$ 15.00	\$ 32280.00	\$ -
Government	Canada	Paseo	None	1817	\$ 10.00	\$ 20.00	\$ 36340.00	\$ -
Government	Germany	Paseo	None	1513	\$ 10.00	\$ 350.00	\$ 529550.00	\$ -
Government	Mexico	Velo	None	1493	\$ 120.00	\$ 7.00	\$ 10451.00	\$ -

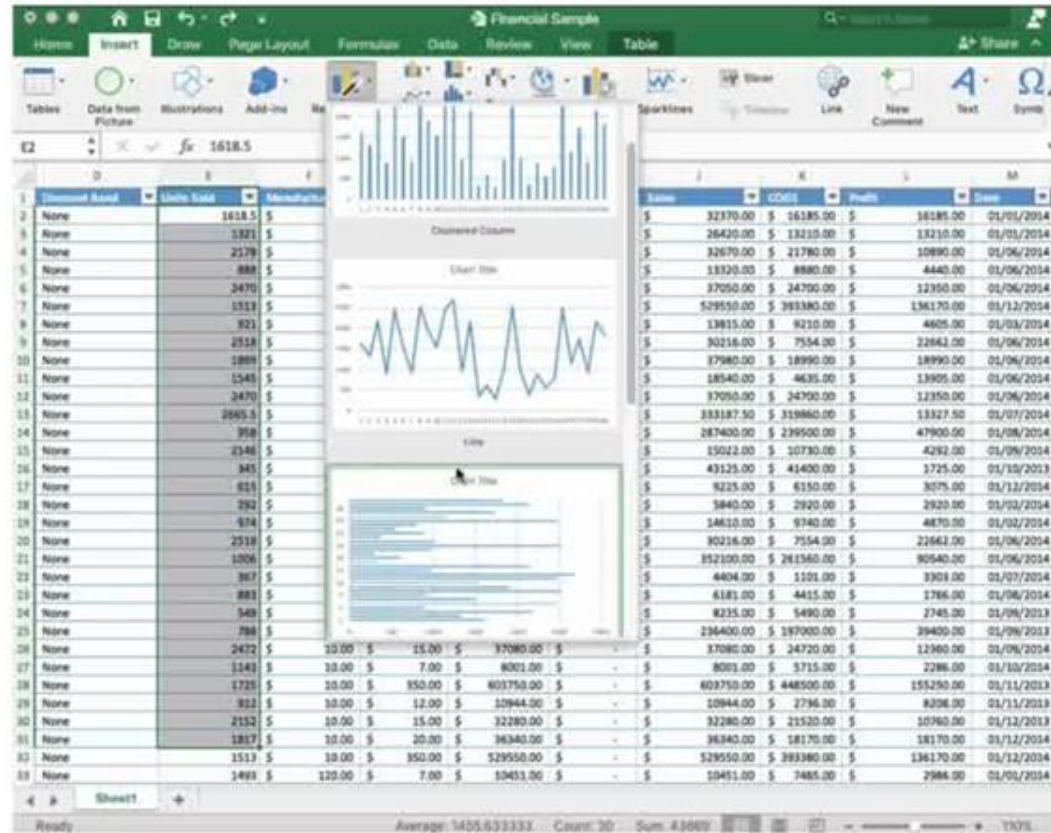
MICROSOFT EXCEL

Company

A big part of the problem

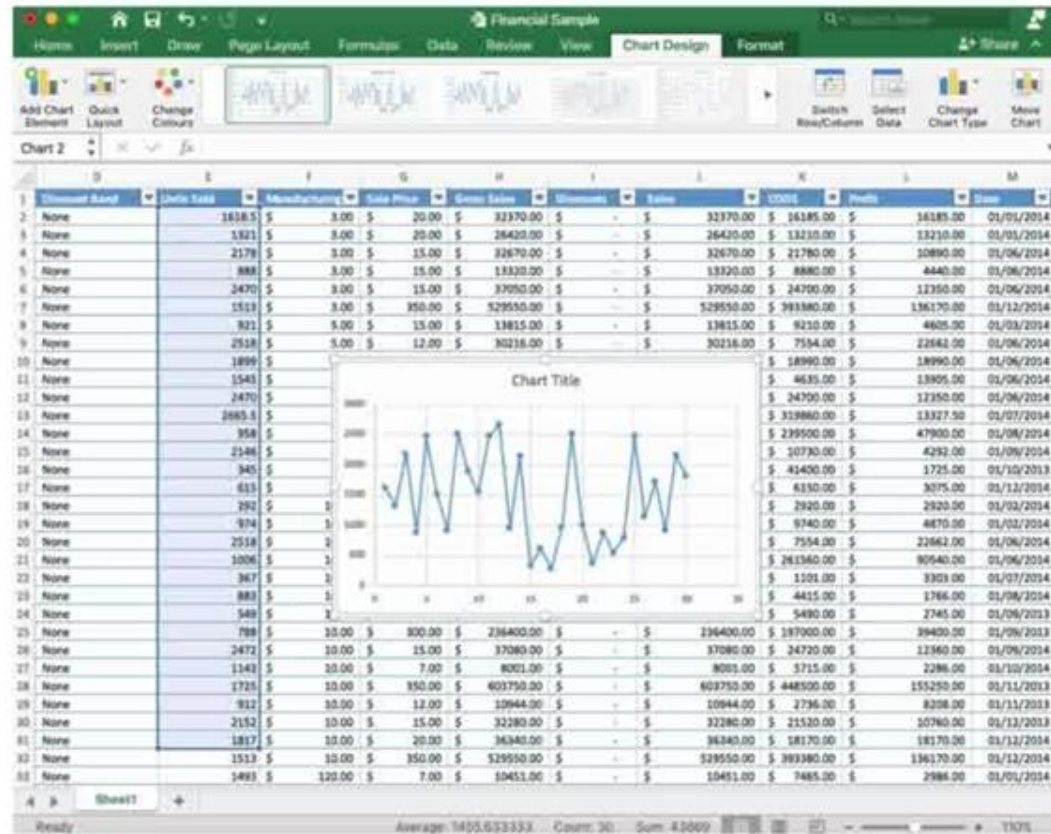
Discount Band	Units Sold	Manufacturing	Sale Price	Gross Sales	Discounts	Sales	COGS	Profits	Date
None	1618.5	\$ 8.00	\$ 20.00	\$ 32370.00	\$ -	\$ 32370.00	\$ 16185.00	\$ 16185.00	01/01/2014
None	1821	\$ 8.00	\$ 20.00	\$ 26420.00	\$ -	\$ 26420.00	\$ 13210.00	\$ 13210.00	01/01/2014
None	2178	\$ 8.00	\$ 15.00	\$ 32670.00	\$ -	\$ 32670.00	\$ 21780.00	\$ 10890.00	01/06/2014
None	888	\$ 3.00	\$ 15.00	\$ 13320.00	\$ -	\$ 13320.00	\$ 8880.00	\$ 4440.00	01/06/2014
None	2470	\$ 3.00	\$ 15.00	\$ 37050.00	\$ -	\$ 37050.00	\$ 24700.00	\$ 12350.00	01/06/2014
None	1511	\$ 3.00	\$ 350.00	\$ 529550.00	\$ -	\$ 529550.00	\$ 393380.00	\$ 136170.00	01/12/2014
None	921	\$ 5.00	\$ 15.00	\$ 13815.00	\$ -	\$ 13815.00	\$ 9210.00	\$ 4605.00	01/01/2014
None	2518	\$ 5.00	\$ 12.00	\$ 30216.00	\$ -	\$ 30216.00	\$ 7554.00	\$ 22662.00	01/06/2014
None	1899	\$ 5.00	\$ 20.00	\$ 37980.00	\$ -	\$ 37980.00	\$ 18990.00	\$ 18990.00	01/06/2014
None	1545	\$ 5.00	\$ 12.00	\$ 18540.00	\$ -	\$ 18540.00	\$ 4635.00	\$ 13905.00	01/06/2014
None	2470	\$ 5.00	\$ 15.00	\$ 37050.00	\$ -	\$ 37050.00	\$ 24700.00	\$ 12350.00	01/06/2014
None	2665.5	\$ 5.00	\$ 125.00	\$ 333187.50	\$ -	\$ 333187.50	\$ 333860.00	\$ 13327.50	01/07/2014
None	358	\$ 3.00	\$ 800.00	\$ 287400.00	\$ -	\$ 287400.00	\$ 239500.00	\$ 47900.00	01/08/2014
None	2148	\$ 3.00	\$ 7.00	\$ 15022.00	\$ -	\$ 15022.00	\$ 10730.00	\$ 4292.00	01/09/2014
None	345	\$ 5.00	\$ 125.00	\$ 43125.00	\$ -	\$ 43125.00	\$ 41400.00	\$ 1725.00	01/10/2013
None	615	\$ 5.00	\$ 15.00	\$ 9225.00	\$ -	\$ 9225.00	\$ 6150.00	\$ 3075.00	01/11/2014
None	292	\$ 10.00	\$ 20.00	\$ 5840.00	\$ -	\$ 5840.00	\$ 2920.00	\$ 2920.00	01/02/2014
None	974	\$ 10.00	\$ 15.00	\$ 14610.00	\$ -	\$ 14610.00	\$ 9740.00	\$ 4870.00	01/02/2014
None	2518	\$ 10.00	\$ 12.00	\$ 30216.00	\$ -	\$ 30216.00	\$ 7554.00	\$ 22662.00	01/06/2014
None	1006	\$ 10.00	\$ 350.00	\$ 352100.00	\$ -	\$ 352100.00	\$ 261560.00	\$ 90540.00	01/06/2014
None	367	\$ 10.00	\$ 12.00	\$ 4404.00	\$ -	\$ 4404.00	\$ 1101.00	\$ 3303.00	01/07/2014
None	883	\$ 10.00	\$ 7.00	\$ 6181.00	\$ -	\$ 6181.00	\$ 4415.00	\$ 1766.00	01/08/2014
None	549	\$ 10.00	\$ 15.00	\$ 8235.00	\$ -	\$ 8235.00	\$ 5490.00	\$ 2745.00	01/09/2013
None	788	\$ 10.00	\$ 300.00	\$ 236400.00	\$ -	\$ 236400.00	\$ 187000.00	\$ 49400.00	01/09/2013
None	2472	\$ 10.00	\$ 15.00	\$ 37080.00	\$ -	\$ 37080.00	\$ 24720.00	\$ 12360.00	01/09/2014
None	1143	\$ 10.00	\$ 7.00	\$ 8001.00	\$ -	\$ 8001.00	\$ 5715.00	\$ 2286.00	01/10/2014
None	1725	\$ 10.00	\$ 350.00	\$ 603750.00	\$ -	\$ 603750.00	\$ 448500.00	\$ 155250.00	01/11/2013
None	812	\$ 10.00	\$ 12.00	\$ 10944.00	\$ -	\$ 10944.00	\$ 2736.00	\$ 8208.00	01/11/2013
None	2152	\$ 10.00	\$ 15.00	\$ 32280.00	\$ -	\$ 32280.00	\$ 21520.00	\$ 10760.00	01/12/2013
None	1817	\$ 10.00	\$ 20.00	\$ 36340.00	\$ -	\$ 36340.00	\$ 18170.00	\$ 18170.00	01/12/2014
None	1513	\$ 10.00	\$ 350.00	\$ 529550.00	\$ -	\$ 529550.00	\$ 393380.00	\$ 136170.00	01/12/2014
None	1488	\$ 120.00	\$ 7.00	\$ 10451.00	\$ -	\$ 10451.00	\$ 7485.00	\$ 2966.00	01/01/2014

A big part of the problem



MICROSOFT EXCEL

A big part of the problem



A big part of the problem

Plotting is done in autopilot mode

Technological achievements make us forget that drawing a plot is a rational, creative process, nothing automatic.

We don't plan our plots ahead, instead we expect technology to give us the solution

We need to put more thinking into it

A screenshot of a text editor window titled "plotWealthDistribution.r — WealthDistribution". The window contains R code for reading a CSV file and creating a bar plot with ggplot2. The code is as follows:

```
1 library(ggplot2)
2
3 d <- read.csv("WealthDistribution.csv", sep=";", header=T, comment.char="#")
4 subd <- data.frame(d$Percentile, d$X2016)
5 names(subd) <- c("Percentile", "Wealth")
6
7 # PLOT 1
8 p <- ggplot(data=subd, aes(x=Percentile, y=Wealth)) +
9   geom_bar(stat="identity", fill="#E68821") +
10  theme_minimal() +
11  theme(legend.title = element_blank(),
12        legend.text = element_text(size = 12))
```

TEXT EDITOR WITH R AND GGPLOT CODE

What can we do?

The Ingredients for Good Data Visualization

Good Data Visualization

Data to plot

The Ingredients for Good Data Visualization

Good Data Visualization

Data to plot

Give a clear message

Design principles

The Ingredients for Good Data Visualization

Good Data Visualization

Data to plot

Give a clear message

Respect the data

Design principles

Statistical concepts

The Ingredients for Good Data Visualization

Good Data Visualization

Data to plot

Give a clear message

Respect the data

Right representation

Design principles

Statistical concepts

Plots

The Ingredients for Good Data Visualization

Good Data Visualization

Data to plot

Give a clear message

Respect the data

Right representation

Design principles

Statistical concepts

Plots

The Ingredients for Good Data Visualization

Good Data Visualization

Data to plot

Give a clear message

Respect the data

Right representation

Appropriate tools

Design principles

Statistical concepts

Plots

ggplot2

