Quantitative methods

Lesson 2

Daróczi Gergely

Corvinus University of Budapest, Hungary

2011 February 15

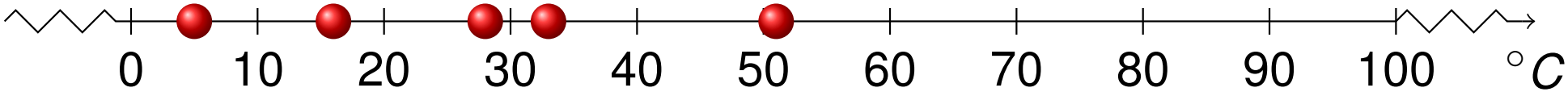
Outline

Measurements

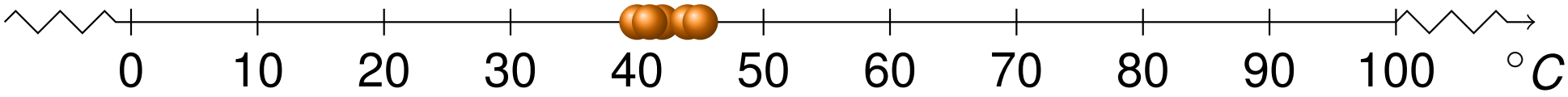
Types of error

Measuring the temperature of a bowl of water kept at 60°C

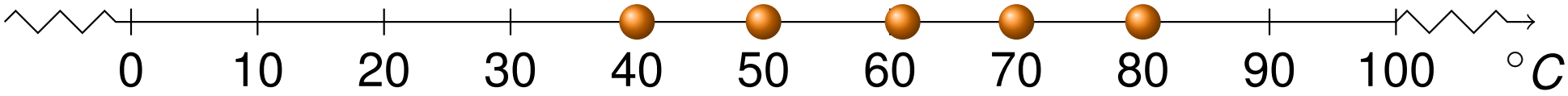
Not reliable, not valid measurements:



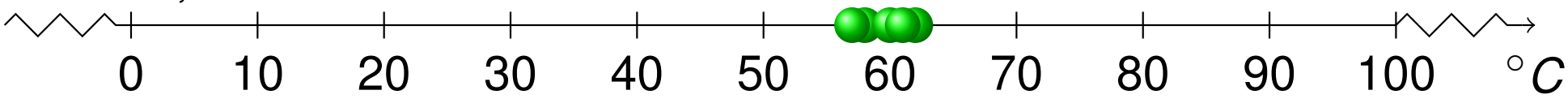
Reliable, not valid measurements:



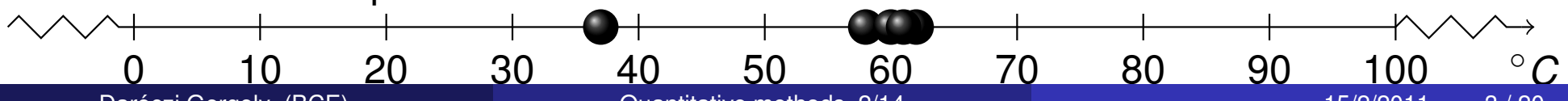
Not reliable, valid measurements:



Reliable, valid measurements:

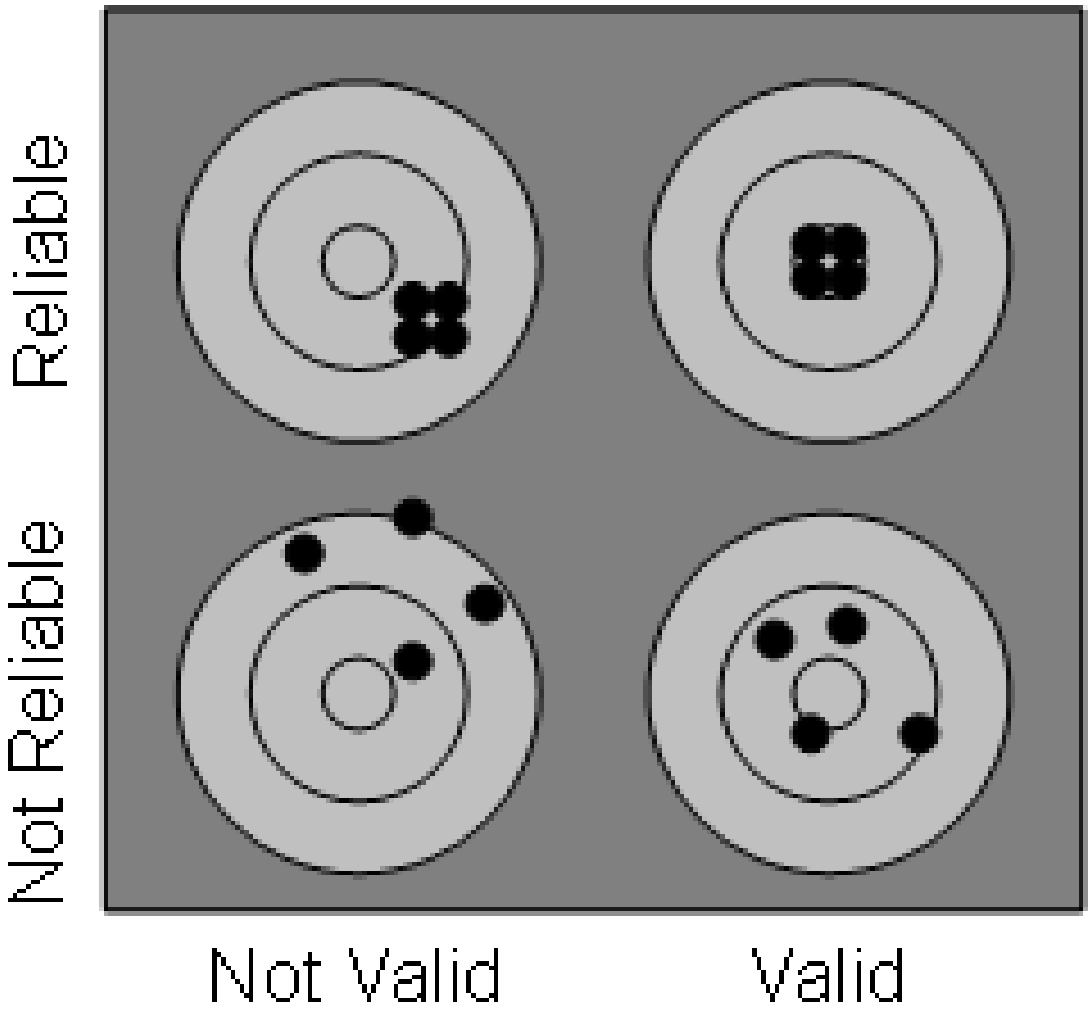


What about this sample?



Gone Hunting

Types of error

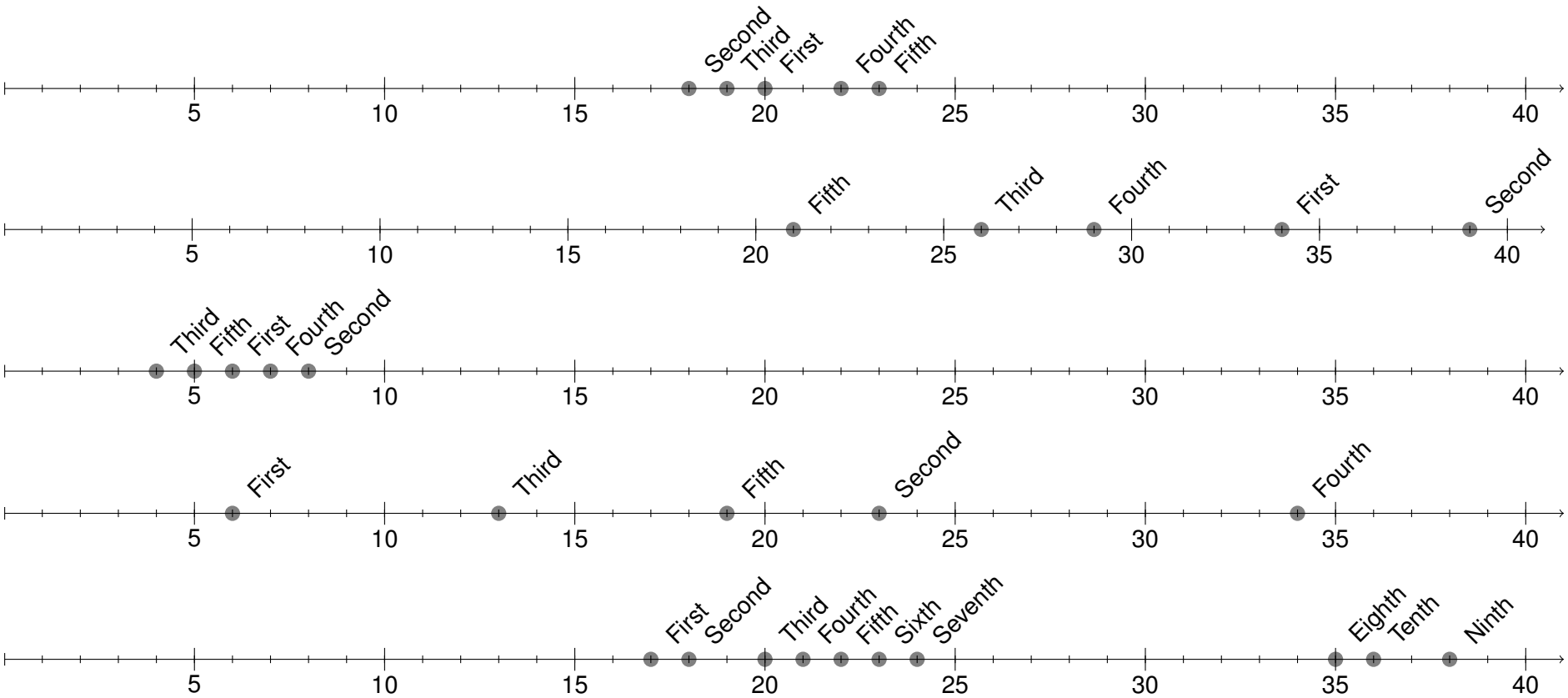


Source: customerthink.com

Test your knowledge!

Types of error

Measuring the temperature of a bowl of water kept at 20°C:



Which set of measures should be considered as valid, reliable and as accurate?

Possible goals of research

Brainstorming

Which of the followings would make a good topic for research?

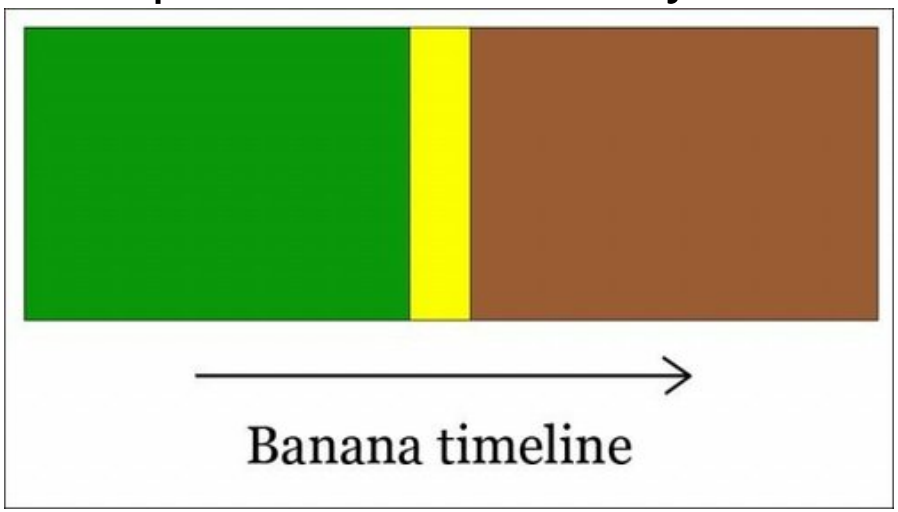
- How many homeless people lives in Budapest?
- What is the most popular way of suicide?
- What color is a banana?

Possible goals of research

Example of topics seeming meanless

Based on a real online survey:

A simple time series analysis:



Possible goals of research

Brainstorming

Which of the followings would make a good topic for research?

- How many homeless people lives in Budapest?
- What is the most popular way of suicide?
- What color is a banana?
- How much does a pile of apples cost?
- How kind of relationship can be found between bodyweight and hair-color?
- What is the best book in the world?

What could be the main factors of choosing a topic?

Definition: exploratory vs. descriptive vs. explanatory study

Units of analysis

Units, cases, objects

What entities are described and compared?

- Are older people more afraid of crime than younger people?
- Does economic development lower the birth rate?
- Which Hungarian soccer team has the utmost fan club?
- The longer the emgagement period, the longer the marriage.
- What is the most prestigious occupation?

Do we always ask the units of analysis?

Units of analysis

Ecological fallacy and so

We found, that 2 million tourists arrived to Budapest in 2010. It was also known, that 2.5 million tourists visited the lake Balaton in the same year.

Could we state that . . .

- the inhabitants of Budapest are not as friendy as people living at the lake Balaton?
- Budapest has less spectacle than the lake Balaton?
- the lake is a bigger tourist goal?
- tourists think that lake Balaton is a lot nicer place than Budapest?
- the beach of Balaton is lot more crowded than the streets of Budapest?

What can be said based on the ebove data?

Variables and attributes

A concrete example

How old are you?

- >18
- 19-24
- 25-30
- 31-40
- 41-100
- 100<

What is variable and what is attribute in the above example?

Variables and attributes

Test your knowledge!

What entities are described and compared?

- Are older people more afraid of crime than younger people?
- Does economic development lower the birth rate?
- Which Hungarian soccer team has the utmost fan club?
- The longer the emgagement period, the longer the marriage.
- What is the most prestigious occupation?

What is variable and what is attribute in the above example?

Types of variables and attributes

Towards methodology

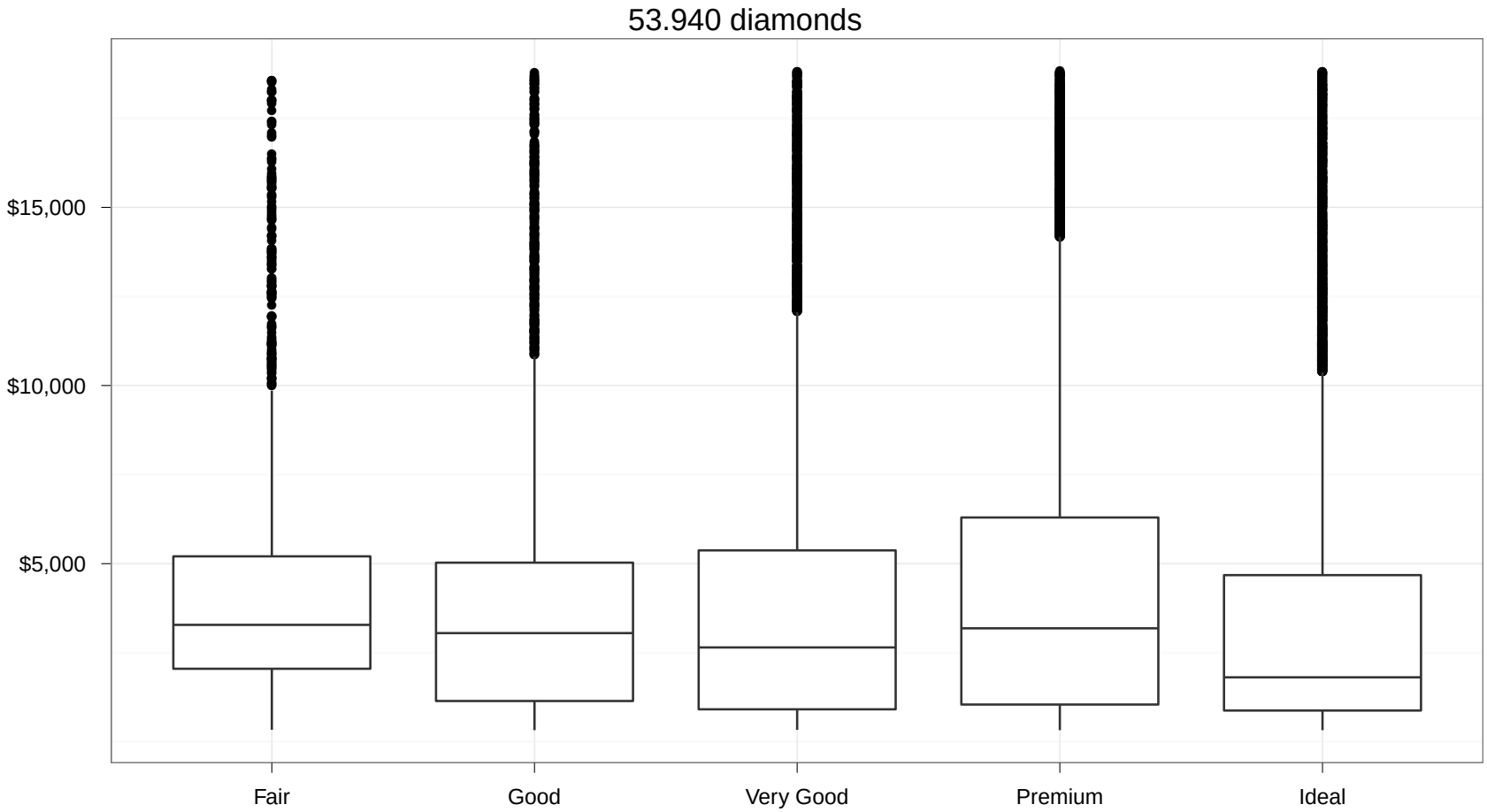
Types of variables

- explanatory variables
  - dependent variables
    - qualitative variables
    - quantitative variables
  - independent variables
    - qualitative variables
    - quantitative variables
- extrameous variables
  - control variables
  - other variables

Let's make up some examples based on the above list!

Relation between variables

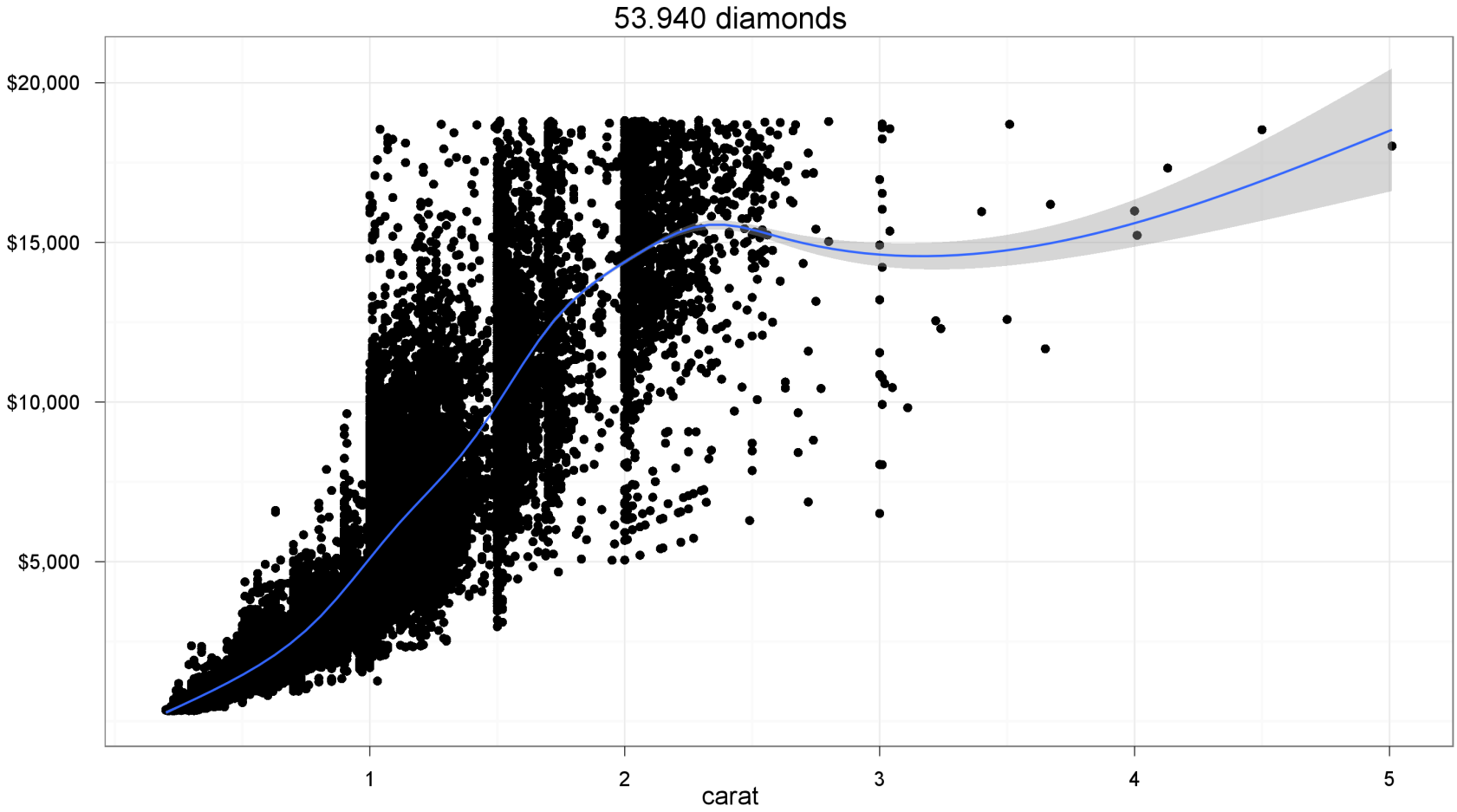
A visual example




```
ggplot(diamonds, aes(cut, price)) + geom_boxplot() + xlab('') + ylab('') +
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

Relation between variables

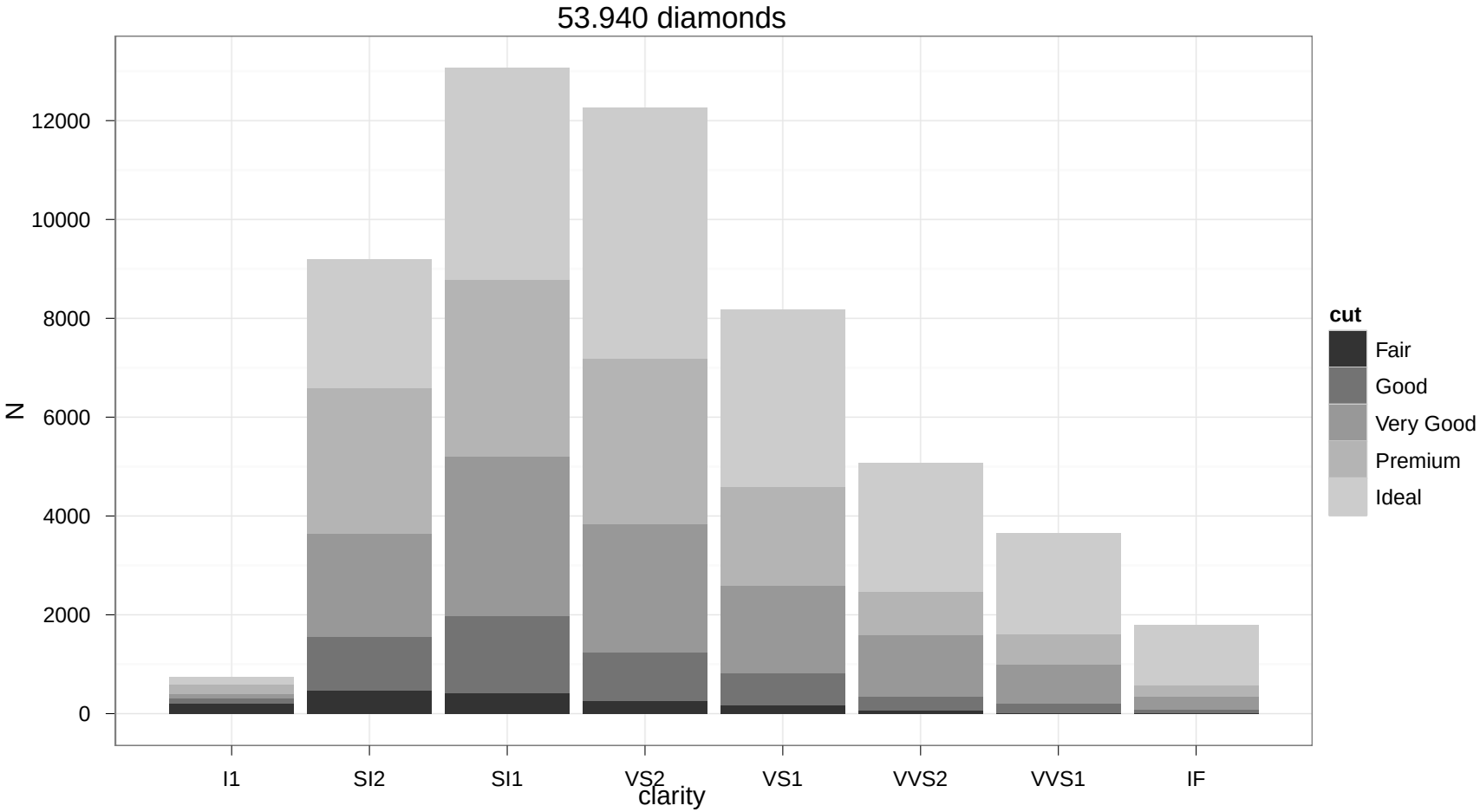
A visual example




```
ggplot(diamonds, aes(carat, price)) + geom_point() + geom_smooth() + ylab('') +
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

Relation between variables

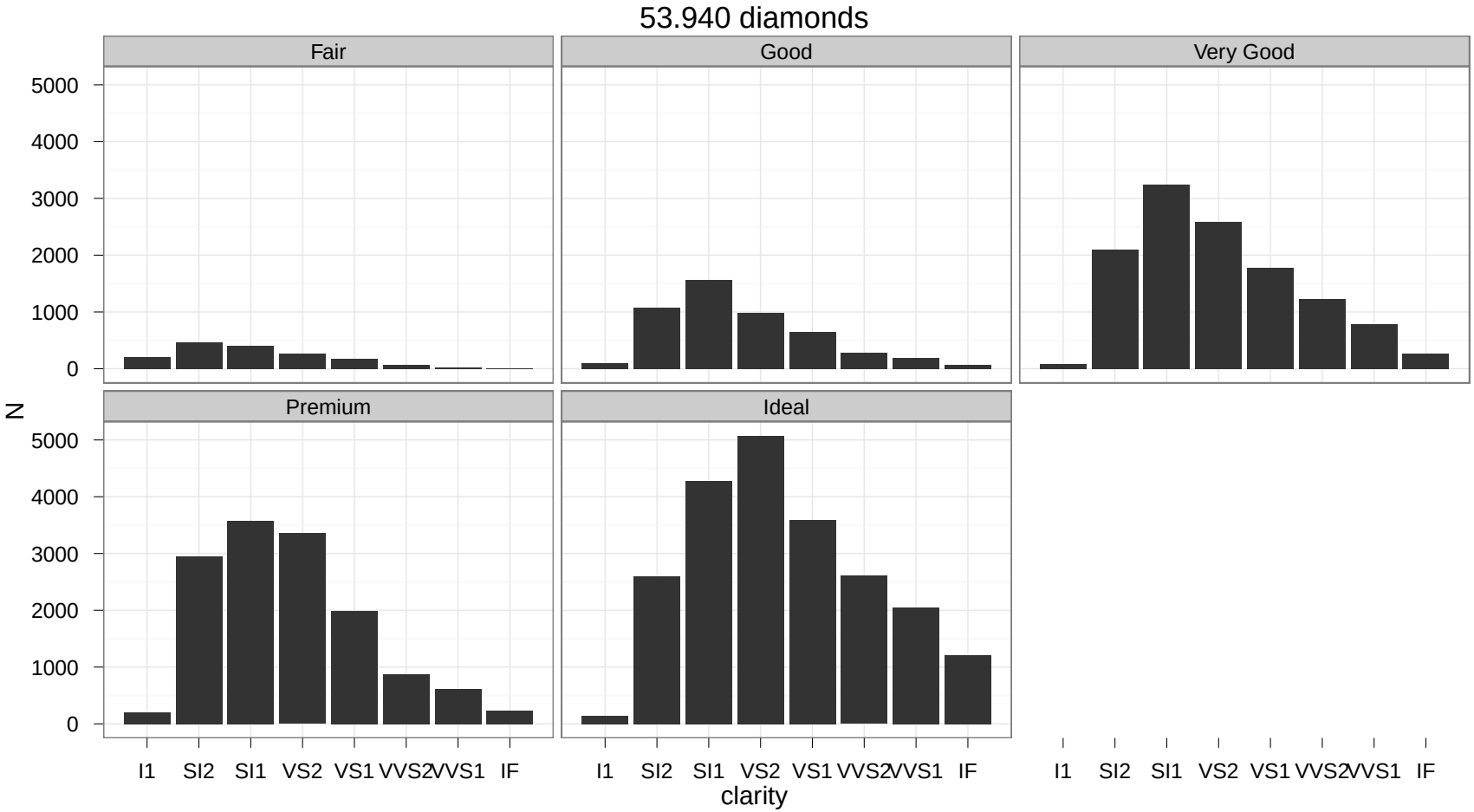
A visual example




```
ggplot(diamonds, aes(clarity, fill=cut)) + geom_bar() + ylab("N") +
theme_bw() + opts(title="53.940 diamonds")
```

Relation between variables

A visual example




```
ggplot(diamonds, aes(clarity)) + geom_bar() + ylab("N") + facet_wrap(~ cut) +
theme_bw() + opts(title="53.940 diamonds")
```

Types of variables and attributes in practice

Relationship between variables

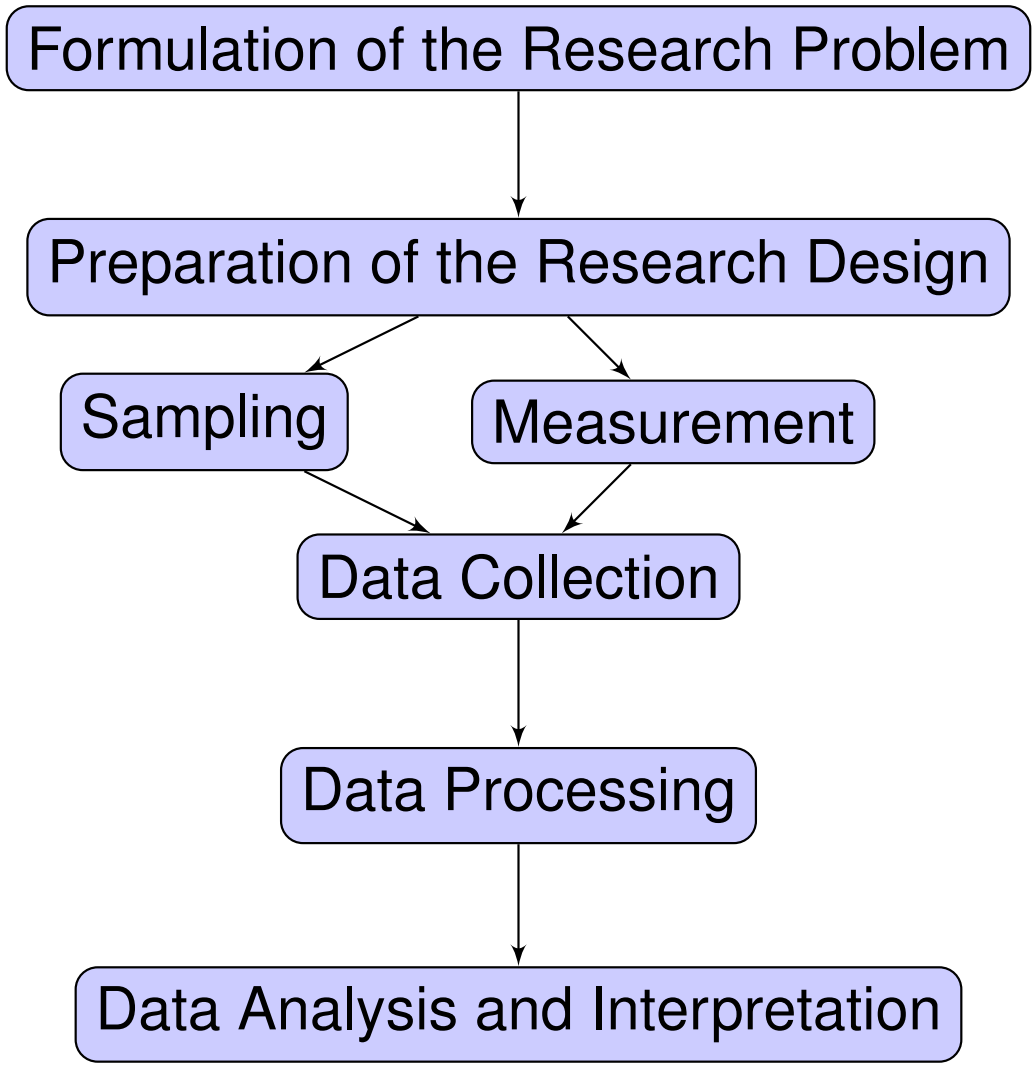
Possible relationship between variables:

- association,
- correlation,
- spurious relationship,
- influence,
- direction of influence,
- **causality**.

What is hypothesis?

Stages of Social Research

A flowchart



It was a pleasure!

Daróczi Gergely
daroczi.gergely@btk.ppke.hu