CEM1002

Neil Montgomery

2015-09-14

Course Basics

- Neil Montgomery
- `neilm@mie.utoronto.ca`
- BA8132 office hours. . . any time
- Official description is misleading (from 2013)
- Course will be primarily "statistics" with cities issues as applications
- I'm working on field trips and/or guest lectures
- Evaluation:
  - 4 assignments each worth 25%
  - An assignment may have multiple parts with different due dates (esp. the first one)
- No required texts. For the basics, any recent "Stats 101" book is fine.

Software

Statistics requires software. In this course we'll use R.

Pros:

- High quality free software environment
- Help desk: Google
- Extensions via rich package eco-system
- Not point-and-click (reproducible workflow)

Cons:

- Not point-and-click (learning curve)

To install:

- Optional but recommended: Install LaTeX (to produce PDF reports) `latex-project.org`
- Install `R` itself: `r-project.org`
- Install Rstudio: `rstudio.com`

Assignment 1 (full details to be on course website)

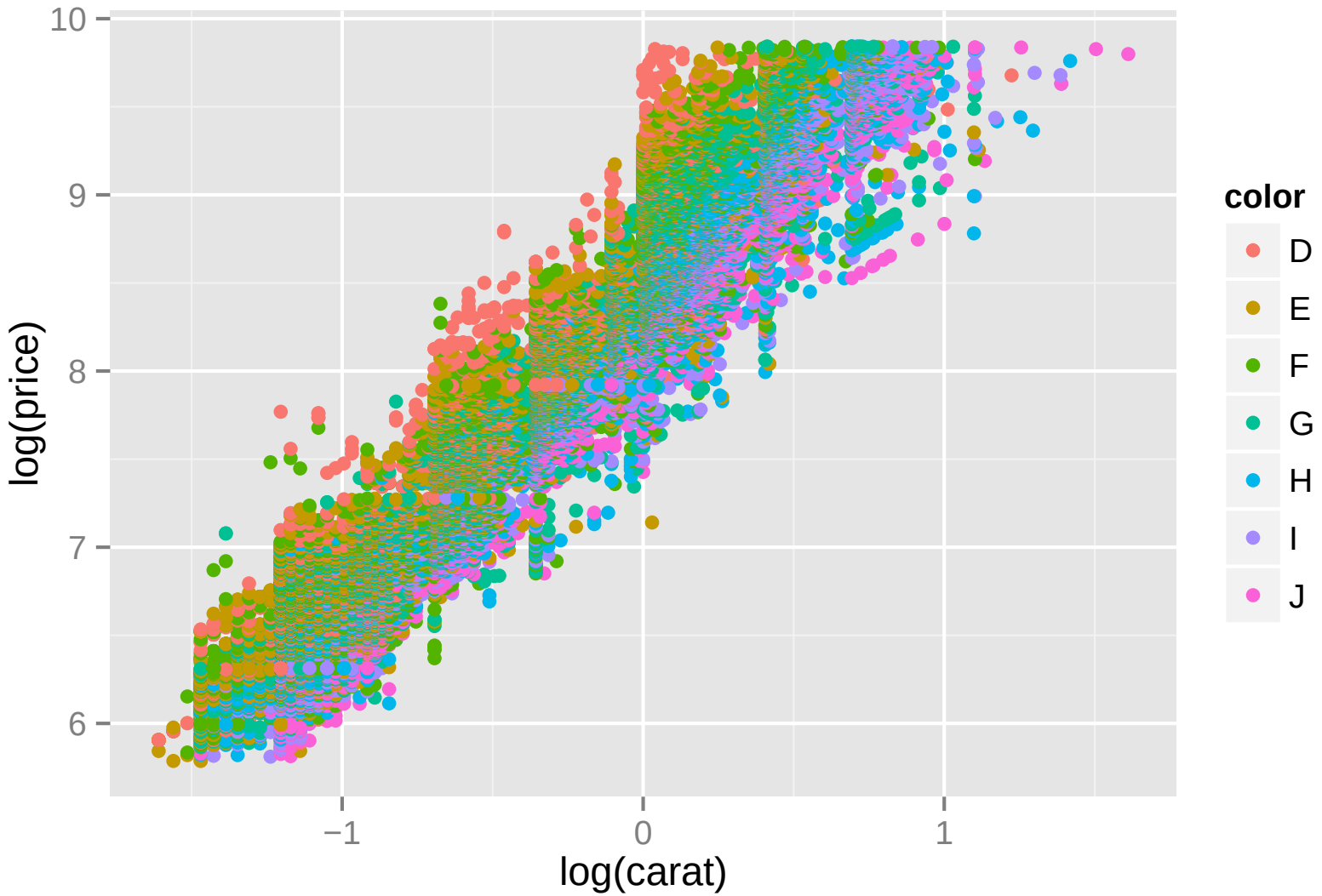
- Find a dataset related to one or more cities and perform a detailed exploratory data analysis (EDA) on it.
- Hint: Google " open data"
- Three deadlines:
  - September ~~21~~ **22**: Install `R` environment, find (interim) dataset(s), and produce a skeleton report with a numerical summary and a graphical summary.
  - September ~~28~~ **29**: Using final dataset(s) provide interim report.
  - October ~~5~~ **9**: Final report.
  - Dates may be adjusted to avoid distaster.

R miscellanea

- Notes and R code used in each class will be available.
- A few packages will get used quite a bit and you'll need to install them (once) and import them (every time). Start with:
  - `dplyr`
  - `ggplot2`
- Rstudio makes good workflow easy with R markdown. Examples:
  - Two plus three equals 5.
  - The `diamonds` example dataset that comes with `ggplot2` has 53940 records in it.

R miscellanea (plot example)

- The plotting package `ggplot2` will seem strange at first but is the gold standard.



Exploratory Data Analysis

- What is a dataset?
- Classification of variable types (not to be taken too seriously)
- Numerical summaries
- Graphical summaries