



# 1-04 Ordered and unordered factors

CSI 500

Course material derived from:

An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics

Version 3.4.3 (2017-11-30)

<https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>

# What are factors?

- R uses "factors" as a way to represent nominal (also called "categorical") values
  - used in model formulas, statistical tests
  - can also be used to simplify data displays and improve readability
- Factors are stored internally as integers
  - makes processing more efficient

# Example: processing survey data

- You have conducted a survey of Data Science students at your university, asking their opinion of the R/Rstudio environment for research.
- Which of these responses most closely matches your opinion of R / Rstudio as a tool for data science research? Choose one response
  - ☐ very low
  - ☐ low
  - ☐ medium
  - ☐ high
  - ☐ very high

# Example: processing survey data

- Let's say you got 20 responses and encoded them in a variable named "resp\_data"
- you make a set of response labels corresponding to your questionnaire instrument called "q\_text"
- you factorize the coded responses to align with the questionnaire text in a variable called "resp\_fac"
- you plot the factored data to visualize the results

```
> resp_data
[1] 3 4 4 3 3 5 4 5 5 3 3 4 3 3 5 1 2 2 1 2

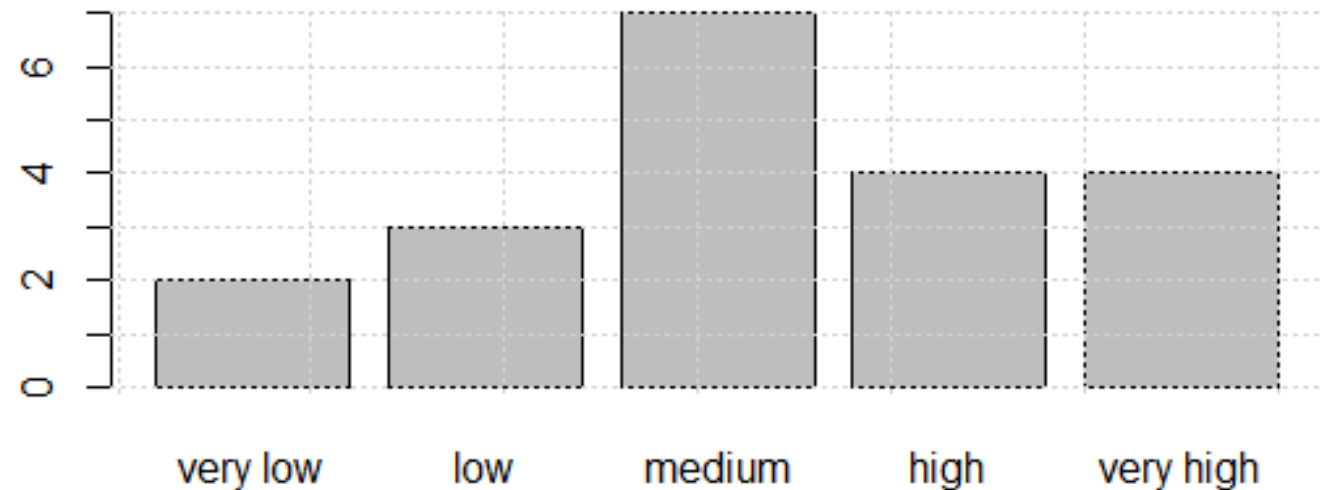
> q_text = c("very_low", "low", "medium", "high", "very_high")
> q_text
[1] "very_low" "low"      "medium"   "high"     "very_high"

> resp_fac = factor(resp_data, labels=q_text, order=TRUE)
> resp_fac
[1] medium  high    high    medium
[5] medium  very_high high    very_high
[9] very_high medium  medium  high
[13] medium  medium  very_high very_low
[17] low     low     very_low low
5 Levels: very_low < low < medium < ... < very_high

> plot(resp_fac)
> grid()
```

# Visualizing the results

- We use a `plot()` command on the factorized results to make a quick barchart
- Not surprisingly, data science students seem to have a high opinion of R/Rstudio



# Summary

- R provides `factors()` as a means of managing categorical or nominal data
  - factors can be ordered if the categories have a natural sort order
  - factors can be unordered if simply nominal values with no natural sort order