Chapter 2 Homework

From Textbook  
Jiawei Han, Micheline Kamber, Jian Pei (2012). Data Mining Concepts and Techniques, 3rd ed. Morgan Kaufmann.

# Page 80, Exercises 2.2, 2.3, 2.4, 2.8

Package for 2.8: “philentropy” and the function for Euclidean and supremum

2.2(a) Mean

> ages <- c(13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30,33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70)

> mean(ages)

[1] 29.96296

2.2(b) Mean  
The most frequently-occurring datum is “25”. It occurs 4 times, hence the length of 4 described below.  
Our data is positively skewed, because the mean is greater than the mode.

> modes(ages)

[,1] [,2]

Value 25 35

Length 4 4

2.2 (c) Midrange is 41.5  
max <- max(ages)  
min <- min(ages)  
midrange <- (min + max) / 2

# MIDUS Project Dataset

3.

[**Midlife in the United States (MIDUS 2), 2004-2006**](https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/4652)**(ICPSR 4652)**  
Ryff, Carol; Almeida, David M.; Ayanian, John; Carr, Deborah S.; Cleary, Paul D.; Coe, Christopher; Davidson, Richard; Krueger, Robert F.; Lachman, Marge E.; Marks, Nadine F.; Mroczek, Daniel K.; Seeman, Teresa; Seltzer, Marsha Mailick; Singer, Burton H.; Sloan, Richard P.; Tun, Patricia A.; Weinstein, Maxine; Williams, David

And optionally for biomarkers

4.

[**Midlife in the United States (MIDUS 2): Biomarker Project, 2004-2009**](https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/29282)**(ICPSR 29282)**  
Ryff, Carol D.; Seeman, Teresa; Weinstein, Maxine