Session 4: Graphics

Li (Sherlly) Xie

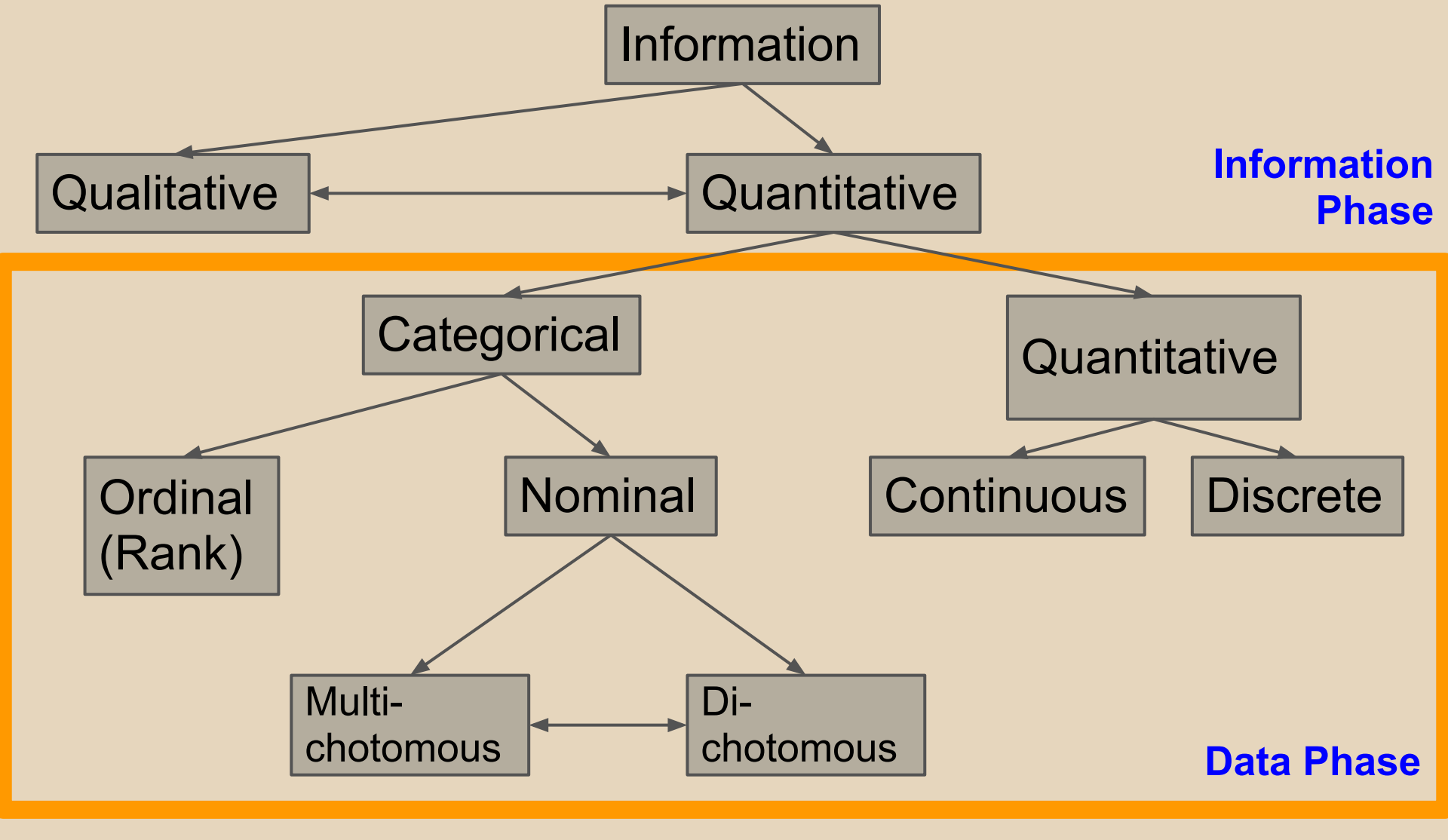
Session 4 Flow

1. Summary of Session 1-3 Materials
    a. Types of Variables
    b. Descriptive Statistics

2. Data Visualization
    a. Univariate/Single-Variable Visualization

Bivariate visualization next week

Sessions 1-3:
From Information to Datum



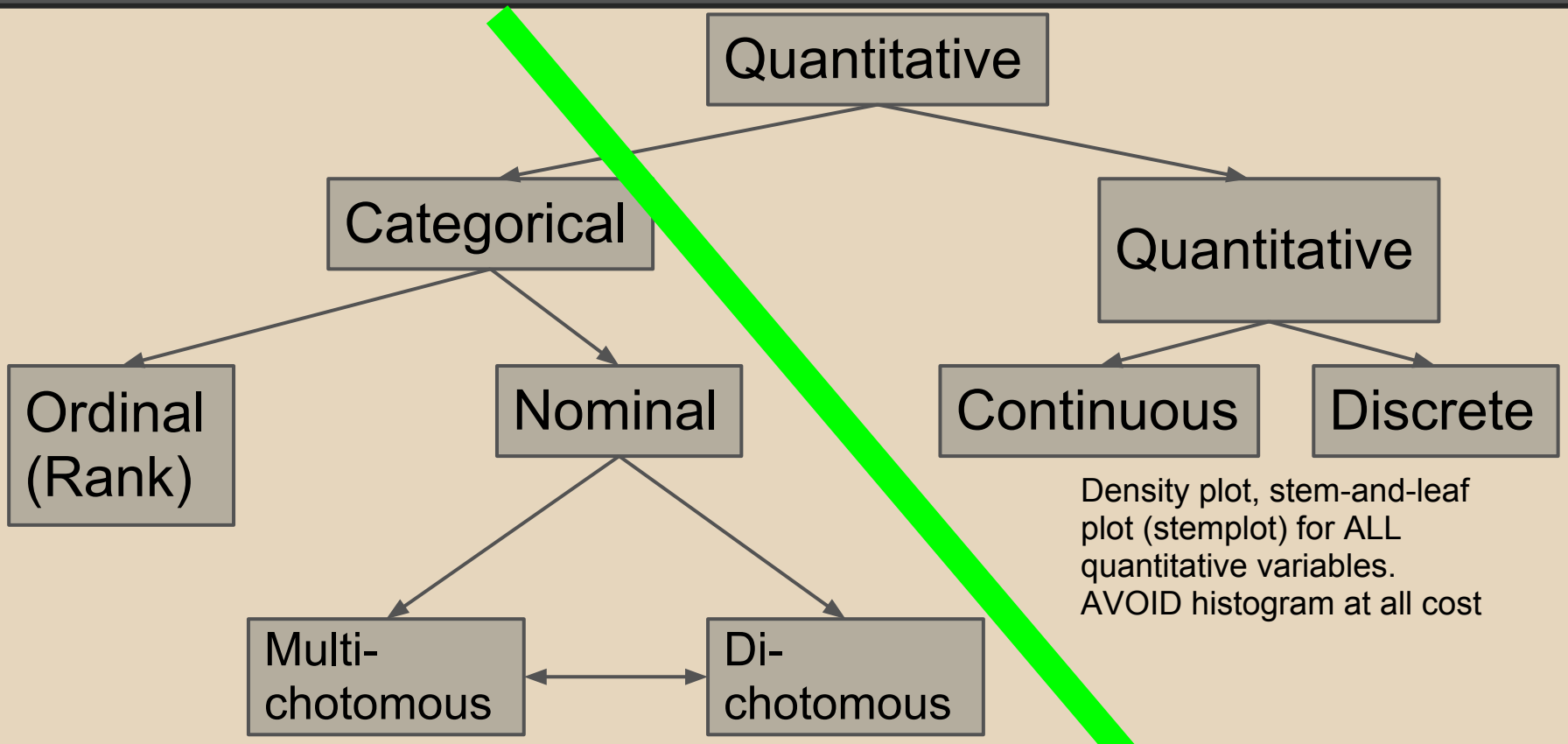
Pop Quiz-Identify The Variable Type

pt_ID	Male	Grade_K12	degree_burn	ER_wait_time.min	visit_accomp_by_parent	Type_of_insur	Date_ER visit	doc_notes	AgeCategory_0.under9_1.age9to18	SBP_admsn	DBP_admsn
33	T	3	1	17	T	CHIP	01/01/2013	multiple burns	0	115	70
1	F	7	1	22	F	PPO	02/07/2013		1	120	75
17	F	11	2	5	T	HMO	01/09/2013	blisters	1	115	80

Categorical: ordinal, nominal (di-/multi-chotomous)

Quantitative: continuous, discrete

Usual Univariate Graphic Displays



Density plot, stem-and-leaf plot (stemplot) for ALL quantitative variables.
AVOID histogram at all cost

Table, barplot (bar graph) for ALL categorical variables.
AVOID pie chart at all cost

Table: Tabulating by Marginal Distribution

Table 1
Descriptive statistics

Variable	n	Minimum	Maximum	Mean	SD
Sex	346	0	1	.51	.50
Family socioeconomic status	343	−1.99	3.44	.02	.96
Maternal BMI	346	16.44	61.99	26.07	6.00
Age 11 pubertal status	322	1.00	4.50	2.13	.81
Age 13 pubertal status	330	1.00	5.00	3.47	.95
Age 15 pubertal status	346	2.00	5.00	4.36	.48
Age 11 mental health symptoms	307	−1.58	3.05	.02	.90
Age 13 mental health symptoms	303	−1.54	4.41	−.00	.85
Age 15 mental health symptoms	292	−1.57	3.46	−.00	.96
Age 11 morning cortisol	297	1.17	2.81	1.22	.02
Age 11 cortisol slope	297	−.02	−.01	−.02	.00
Age 13 morning cortisol	306	1.16	1.38	1.23	.03
Age 13 cortisol slope	306	−.02	−.01	−.02	.00
Age 15 morning cortisol	272	1.17	1.34	1.22	.02
Age 15 cortisol slope	272	−.02	−.01	−.01	.01
Longitudinal morning cortisol	346	1.12	1.20	1.14	.01
Longitudinal cortisol slope	346	−.04	−.01	−.02	.00
Age 11 BMI	325	−2.36	2.81	.34	1.03
Age 13 BMI	331	−2.38	2.69	.42	1.01
Age 15 BMI	319	−3.46	2.64	.42	.97
Age 18 BMI	305	−2.64	2.63	.31	1.05

Adolescent BMI represents *z*-score for age and sex.
Sex is coded 0 = males, 1 = females; BMI = body mass index; cortisol = predicted logged values (μg/dL).

Ruttle PL, Javaras KN, Klein MH, Armstrong JM, Burk LR, Essex MJ. "Concurrent and Longitudinal Associations Between Diurnal Cortisol and Body Mass Index Across Adolescence." J Ado Health. In Press.

Table: Tabulating by Subgroup

Table 1. Demographic data among all levels of body weights

Variables	*Underweight* (n = *52*)	*Normal* (n = *56*)	*Overweight* (n = *38*)	*Obesity* (n = *51*)
Age, years	7.5 ± 4.2 (2.0 to 18.0)	8.3 ± 4.1 (2.0 to 18.0)	8.2 ± 4.2 (2.0 to 18.0)	8.6 ± 4.2 (2.0 to 16.0)
Gender (male)	36 (69.2)	36 (64.3)	29 (76.3)	41 (80.4)
BMI *z* score[†‡§‖ƒ∫]	−1.01 ± 0.30 (−1.5 to 0.0)	−0.31 ± 0.45 (−0.9 to 0.7)	0.23 ± 0.48 (−0.4 to 1.3)	1.20 ± 0.88 (−0.1 to 4.0)
Obstructive[§ƒ∫]	2.58 ± 4.68 (0.0 to 20.4)	0.99 ± 1.29 (0.0 to 4.4)	2.52 ± 5.42 (0.0 to 31.6)	5.80 ± 8.14 (0.0 to 26.3)
Central	0.17 ± 0.35 (0.0 to 1.5)	0.19 ± 0.46 (0.0 to 3.1)	0.17 ± 0.25 (0.0 to 0.8)	0.17 ± 0.56 (0.0 to 3.5)
Mixed	0.07 ± 0.16 (0.0 to 0.9)	0.07 ± 0.16 (0.0 to 0.8)	0.03 ± 0.10 (0.0 to 0.5)	0.01 ± 0.06 (0.0 to 0.3)
Hypopnea[ƒ]	3.54 ± 4.22 (0.0 to 18.2)	2.00 ± 1.55 (0.0 to 6.0)	2.66 ± 2.88 (0.0 to 14.5)	5.02 ± 6.88 (0.0 to 31.8)
AHI[§ƒ∫]	6.4 ± 8.1 (0.0 to 33.9)	3.2 ± 2.1 (0.0 to 8.1)	5.4 ± 6.3 (0.0 to 34.6)	11.0 ± 12.1 (0.0 to 45.8)
AHI (log)[ƒ]	0.51 ± 0.56 (−0.7 to 1.5)	0.44 ± 0.33 (−0.5 to 0.9)	0.54 ± 0.44 (−0.5 to 1.5)	0.74 ± 0.62 (−0.7 to 1.7)
$MinSaO_2$[§ƒ]	87.4 ± 5.6 (76.0 to 96.0)	88.2 ± 5.9 (61.0 to 96.0)	85.3 ± 9.0 (50.0 to 95.0)	82.7 ± 10.0 (50.0 to 95.0)
$MinSaO_2$ (log)[§ƒ]	1.94 ± 0.03 (1.9 to 2.0)	1.94 ± 0.03 (1.8 to 2.0)	1.93 ± 0.05 (1.7 to 2.0)	1.91 ± 0.06 (1.7 to 2.0)
Arousal index	4.6 ± 3.2 (0.8 to 20.5)	4.0 ± 2.3 (1.4 to 12.8)	3.7 ± 2.9 (1.0 to 17.3)	4.2 ± 3.1 (0.4 to 14.7)
Tonsil (⩾3)	33 (75.0)	35 (70.0)	26 (81.3)	29 (64.4)
Adenoid (>0.67)	13 (37.1)	22 (48.9)	20 (62.5)	18 (50.0)
Allergy	33 (63.5)	42 (75.0)	25 (65.8)	29 (56.9)
Sinusitis	6 (11.5)	12 (21.4)	5 (13.2)	8 (15.7)
Asthma	2 (3.8)	2 (3.6)	1 (2.6)	2 (3.9)

Abbreviations: AHI, apnea–hypopnea index; BMI, body mass index; $MinSaO_2$, minimum arterial oxygen saturation. *Values are given as mean ± /s.d., mean ± / s.d. (range), or No (%) unless otherwise indicated. [†]Significant difference using Bonferroni posteriori comparisons ($P<0.05$; underweight vs normal). [‡]Significant difference using Bonferroni posteriori comparisons ($P<0.05$; underweight vs overweight). [§]Significant difference using Bonferroni posteriori comparisons ($P<0.05$; underweight vs obesity). [‖]Significant difference using Bonferroni posteriori comparisons ($P<0.05$; normal vs overweight). [ƒ]Significant difference using Bonferroni posteriori comparisons ($P<0.05$; normal vs obesity). [∫]Significant difference using Bonferroni posteriori comparisons ($P<0.05$; overweight vs obesity).

Kang KT, Lee PL, Weng WC, Hsu WC. "Body weight status and obstructive sleep apnea in children". Int J Obesity 2012(36):920-4.

Bar graph

Purpose: Show RELATIVE frequency or proportion in each category.

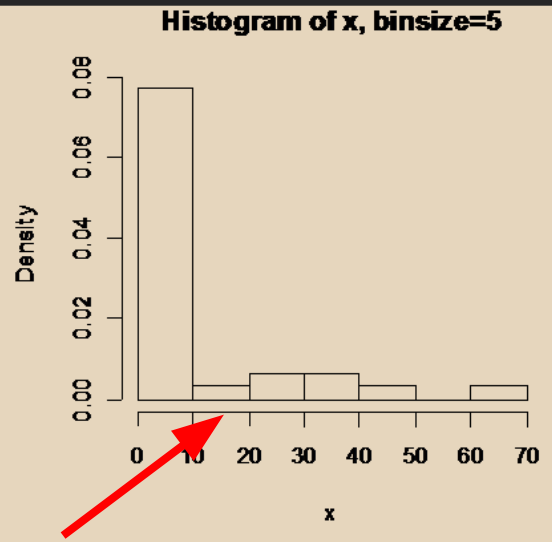
If showing proportions, All bars add up to 100%;

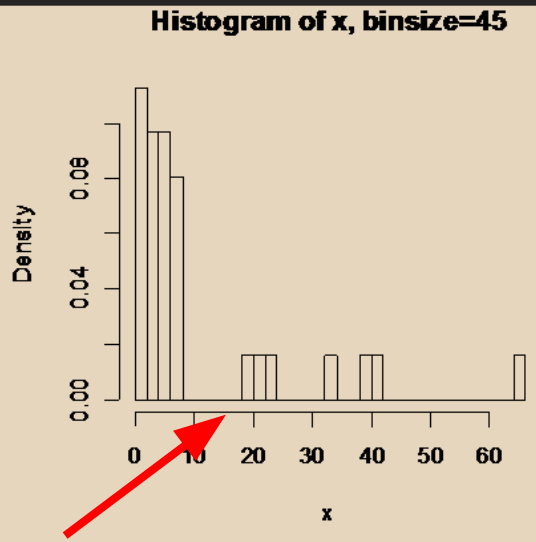
If showing frequencies (counts), all counts add up to the total sample size;

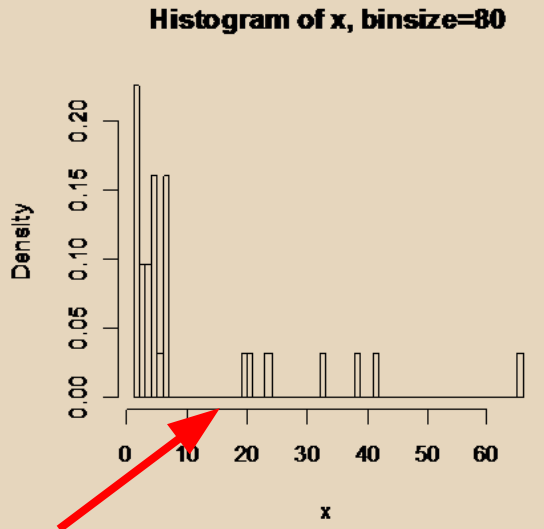
Variants:

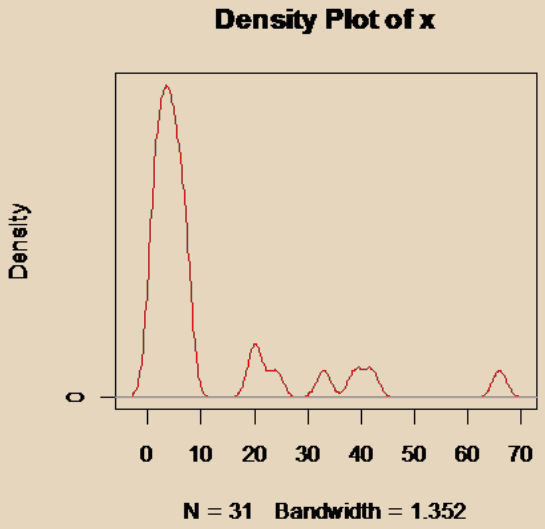
Stacked, clustered, stacked and clustered

The Danger of Histogram: Appearance under influence of bin size

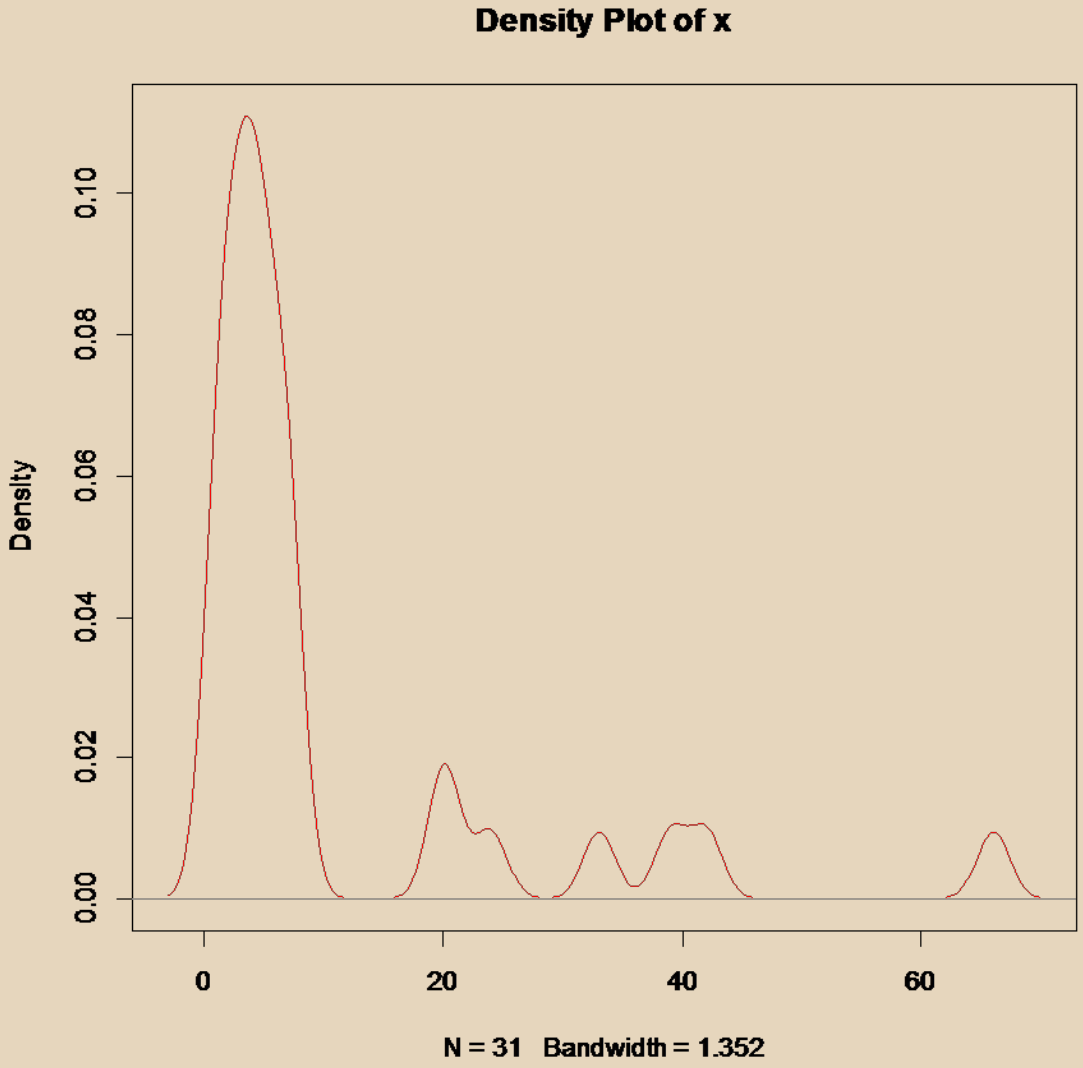








Recommended: Density Plot



A closer look at the empirical distribution of x

Handling Skewed Distributions of Variables in the Data Set

WITH CAUTION, take out outlying values or transform. In either case,

1. Meticulously describe the distribution with/without transformation (outlying values)

2. Make appropriate assumptions about the transformations and/or outlying values

3. Act and justify your actions, NOT with probability/statistics but with science/medicine

Density Plot

The amount being plotted is an approximation to the probability density function of the population from which your data is drawn.

Recommended: download Excel macro "kernel zip" at

http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/kerneldensities.asp

Original article

http://www.rsc.org/images/data-distributions-kernel-density-technical-brief-4_tcm18-214836.pdf

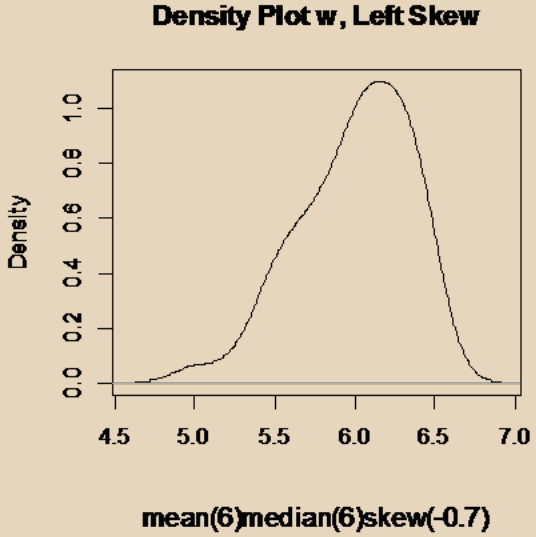
For good karma, please drop a thank-you note at

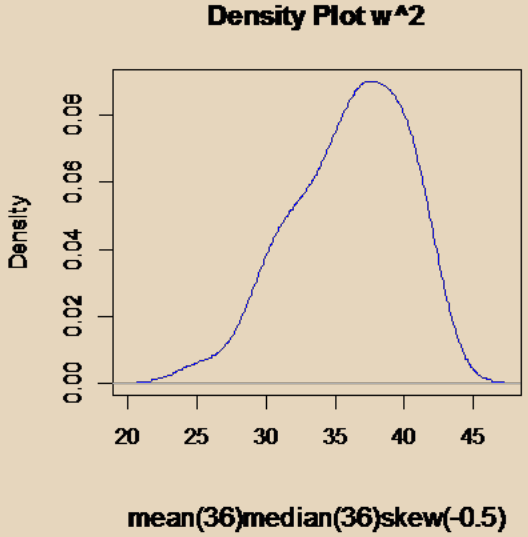
http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/Contacts.asp

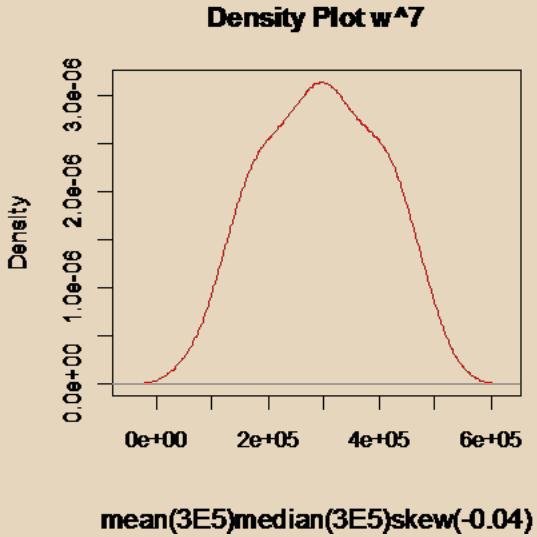
if you decide to use the Excel macro

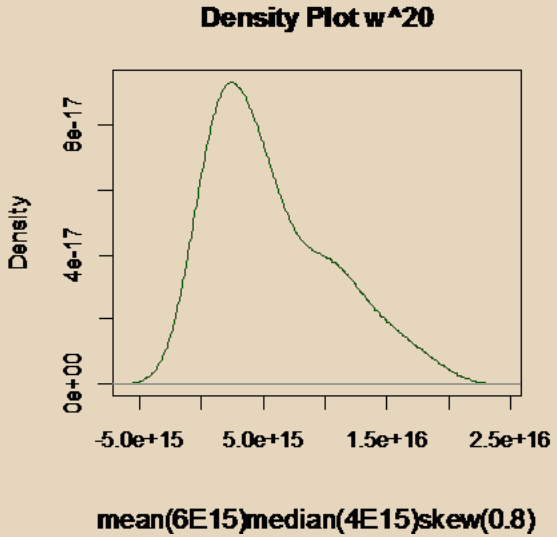
SPSS: http://www.ats.ucla.edu/stat/spss/library/ggraph_examples.htm

Proper Transformations of data whose distribution is left-skewed



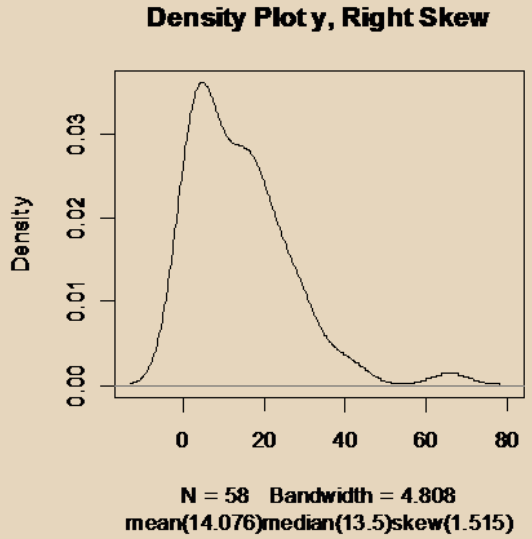


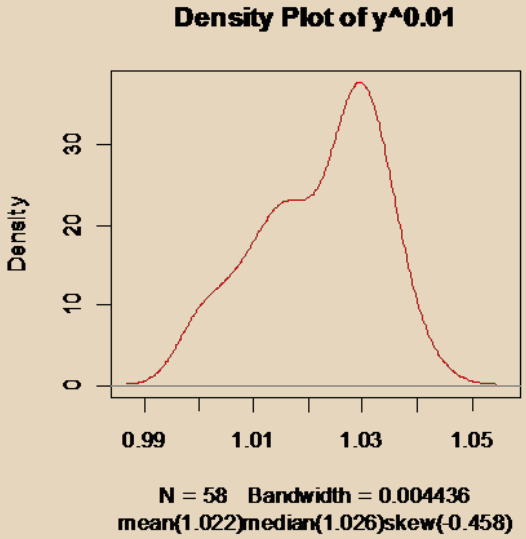


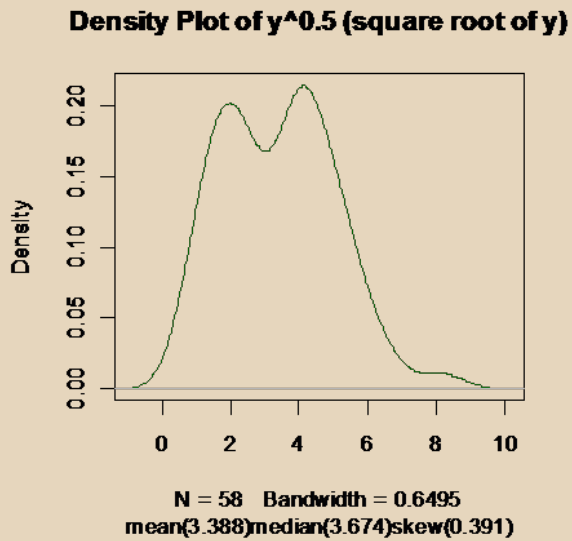


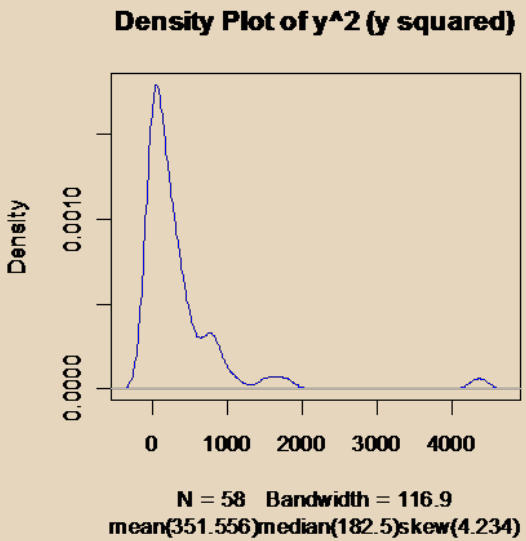
Cannot eyeball everything. Transform, calculate skewness, transform again, calculate again...

Proper Transformations of data whose distribution are right-skewed









*Check for the skewness when comparing transformations. One is better if it gives skewness closer to 0 AND results in a smoother distribution

*Logarithmic transformations are often used to transform right-skewed variable distributions as well

Stem and Leaf Plot

Excel macro: statistics.unl.edu/faculty/bilder/stat2023/excel/stem-and-leaf.xls

Again, for good karma:

bilder@unl.edu; chris@chrisbilder.com

SPSS: Analyze -> Descriptive Statistics -> Explore -> Plots -> Stem-and-Leaf

http://www.math.ou.edu/~mcknight/4753/spss/SPSS4.pdf