Oluwoakami Omotonde
DATA 6666 Homework #1

1.8 Smoking habits of UK residents
a) Each row represents each participant in the study.
b) There are 1691 participants total
c) sex: categorical
age: discrete numerical (we don't have 45½)
marital status: categorical
gross income: categorical, ordinal
smoke: categorical
amt wkends: discrete numerical
amt wkdays: discrete numerical

1.10 Cheaters, Scope of inference
a) population: children between ages 5 and 15
sample: 1160 children ages 5-15
b) The study cannot be used to generalize the population. The sample is too small to allow establishment of causal relationships. Also, there are other factor that could play into results; no control group.

1.28 Reading the paper.
a) I do not think we can ~~think~~ conclude that smoking causes dementia. Of 23,123 participants 25% had dementia. This is not a huge amount to start with. There is some correlation but not causation.
b) No, the statement is not justified. There is a correlation between students w/ behavioral issues and bullies w/ sleep disorders.

1.36 Exercise and Mental health
a) This is an experiment.
b) control: group instructed not to exercise
treatment: group instructed to exercise 2x a week.

1 b) Q1 - below $300000 (25%)
Q2 - below $600,000 (50%) ⇒ median
Q3 below $900,000 (75%)
very few more than $1,200,000
distribution would be symmetric. mean is best representation of typical observation. IQR = $600,000 = 900,000 - 300000. In this case I think we can use IQR or S.D.

c) left skewed distribution, median, IQR since it isn't affected by "few excessive drinkers".

d) left skewed, median, and IQR isn't affected by the much higher salaries.

1.70 Heart Transplants
There are more patients who got a transplant and are alive, therefore, survival is dependent on transplant.

b) boxplot shows survival time higher for treatment group so I would deduce it is effective.

c) treatment group: 1/3 from mosaic
control group: about 1/9 from mosaic

d) i) whether a heart transplant increase lifespan
ii) [redacted], not sure exactly what's being asked.

iii) it suggests that heart transplants has an effect on survival rates.

c) Yes, blocking is used. The blocking variable is age.
d) NO, no blinding is used.
e) This study can be used to establish a causal relationship although if it included a range of how many day per week exercise was done, it would be better. We don't know the sample size, length of time so we can't say if it can be generalized.

1.48 Stats scores : attatched

1.50 Mix and Match
a) Figure a is normally distributed, unimodal and correlates to boxplot # 2.
b) Figure b is a symmetric, uniform distribution and correlates to boxplot #3
c) Figure c is ~~left skewed~~ right skewed and correlated to boxplot #1.

1.56 Distributions and appropriate statistics.
a) Q1 houses below $350000 (25%)
Q2 houses below $450,000 (50%) ⇒ median
Q3 houses below $1000000 (75%)
$6,000,000 (meaningful #)
Because of the meaningful # of $6000000 houses our distribution will be right skewed. Median would be more representative of a typical observation (mean would be higher because of $6000000) IQR would be better used since the meaningful # of $6000000 houses would skew our S.D.