1. Graphics

a) X – position = X – coordinate   } This is
    Y – position = Y – coordinate   } a map!

    size   = size of tree
    shape   = species
    facet   = year
    geom   = point

b) Comparing size of trees across
   years.
   i.e. growth of trees

2. a) money each student has in $

b) Count or # of students

c) That ~~$~~ 42 students
   have between $10 & $20.

d) 1st bar : 60 students have
           between $0 & $10
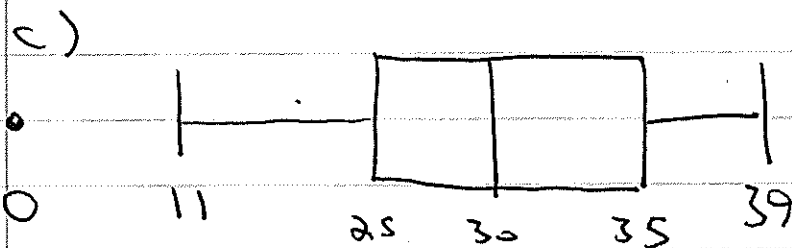   2nd bar : 42 students have
           between $10 & $20.

So 70th student, ie. middle
student, has between $10 & $20

3. House prices

a) 3rd quartile − 1st quartile

$= 35 \quad\quad - 25$

$= 10 = $ width of boxplot

b) Spread. It indicates the spread/variation of the middle 50% of values

c)

Max whisker length $= 1.5 \times IQR$

$= 1.5 \times 10$

$= 15$

Note whiskers extend to observed values 11 & 39
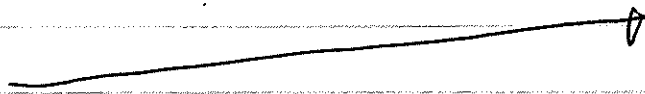
4. a) Short answer

Boxplot:
(without
whiskers)

width of box = IQR = 0

So 25th percentile = median = 75th percentile
= 17

b) hate crimes has 50+1 = 51
rows (for each state + DC)
Warning message says 1
row missing.

So number of points = 51-1
= 50

c)

d) ① Numerical & categorical
② Both numerical
③ " "
④ Numerical & categorical.

Since scatterplots show
relationship between 2 numerical,
② & ④

5. Data Sets.

a) Hourly weather data for each NYC airport. So we need.

Origin, year, month, day, hour

3 airports

Since this is always 2013, we can technically leave this out

need this to ID which date in question.
Ex: March 1st

b) $3 \times 365 \times 24$

\# of airports

\# of days in 2013

\# of hours per day.

1. EDA

a) take africa data frame then
   group by priming then
   summarize with median of countries

b) take africa
   filter year == 2022 then
   select year, priming, countries

c) take africa then
   arrange by countries
   (default ascending order)

2. EDA

a) Both numerical $\Rightarrow$ Scatterplot

b) Variation = spread.
   In a boxplot, length of
   box measures spread
   $= IQR$

   14 countries = 35 - 20 = 15
   94 " = 92 - 38 = 54

c) AAA = height
BBB = how_many_countries
CCC = point
DDD = smooth

d) B/c there are only 4-5 possible values of year, so we would have only 4-5 boxplots side-by-side

4. a) Math = # of countries guessed for someone of height 0. NO PRACTICAL MEANING

b) For every increase of 1 inch in height, there is an associated ~~an~~ decrease of on average 1.86 countries guessed.

c) $\hat{y} = 162 - 1.86 \times x$
$= 162 - 1.86 \times 70$
$= 31.8$.
residual $= y - \hat{y} = 36 - 31.8 = 4.2$

d) Look @ page 5, part c),
scatterplot of countries vs
height.
Smallest height is 58 inches

54 inches is way outside
these points. We are
extrapolating too far.

e) That there is _NO_ relationship
between height _&_ # of
countries guessed.

f) <u>IMO</u> No.

Yes $-1.06 \neq 0$, but that's
probably due to chance.
<u>IMO</u> the slope is not

different enough from 0
to declare there <u>is</u> a

relationship between # of
countries guessed, & height.