

36-315: Statistical Graphics and Visualization

Lab 3

Date: January 28, 2003

Due: end of lab

People are often attracted to cities by the promise of higher income. In this lab, you will test this assumption by comparing per-capita income in Pennsylvania's cities versus the outlying areas. Interspersed throughout this lab are some useful thought questions. You will be asked about them at check-off.

1. Download the files for this lab from the course web page.
2. Open a Word document to record your work.

Start R

3. Start -> Programs -> Class software -> R 1.5.1
4. Set the working directory to My Documents:

File -> Change dir...

5. Load the special functions for this lab:

```
source("lab3.r")
```

Load the data

6. `frame = read.csv("lab3.csv")`
7. Extract the column PCI from `frame`:

```
x = frame[, "PCI"]
```

The data is now a vector of numbers called `x`. It is the per-capita income in each census tract in Pennsylvania, according to the 1990 census.

8. Divide the data according to population density:

```
f = cut.quantile(frame[, "POPPSQMI"], 2, "pop.density")  
s = split(x, f)
```

The two groups are now contained in `s` (a list of two vectors).

Make plots

9. Make boxplots of the two groups:

```
boxplot(s)
```

(Save your plots as you go.) *Can you tell any difference between low and high population density?*

10. Now overlay the full distributions using `density.curve`. *Does any difference stand out now?*
11. To emphasize the difference, make a `density.diff` plot. Compare it to a `density.ditch` plot and a `histodot.ditch` plot. *Which one portrays the differences best?*

By the way, in this plot CMU would be considered “high pop.density” and has PCI = 26,205.

Q-Q plots

12. Using the handout on `qqplot`, make a Q-Q plot of high-density incomes versus low-density incomes. Alternatively, you can use the function `ez.qqplot`, created specifically for this lab:

```
ez.qqplot(s)
```

According to this plot, what transformation relates the incomes in high density areas to the incomes in low density areas?

13. Now extract the column `PCTFEMAL` from `frame`. It is the percentage of women in each tract. Split it by population density, make a Q-Q plot and compare to overlaying the density curves. *What transformation relates the female-percentage in high density areas to low density areas?*
14. Show us your graphs and answer this question: what is the essential difference between the incomes in urban (high density) versus rural (low density) areas in Pennsylvania?