Note: For this midterm exam, you are not allowed to receive help from anyone except me on the exams. For example, you may not talk to other students about the exam problems, and you may not look at other students’ exams. Violations of this policy may result in a 0 on the exam, an F for the course, and/or punishment by the USC Office of Academic Integrity.

1. You are working as a statistical assistant for the United Nations, helping with some statistical analysis to use in a magazine publication that will describe aspects of a sample of UN member countries. The UN has gathered data on 77 countries, measuring 15 variables on each, which represent various characteristics of the nations: LandArea (in square km), Population (in millions of people), Energy (energy usage), Rural (percent of people living in rural areas), Military (spending on the military as a percentage of budget), Health (health care spending as a percentage of GDP), HIV (percent of adults who are HIV positive), Internet (percent of people with Internet access), BirthRate, ElderlyPop (percent of population that is elderly), LifeExpectancy (in years), CO2 (carbon dioxide production), GDP (gross domestic product), Cell (a measure of cell phone usage), Electricity (a measure of consumption of electricity).

The questions that the Secretary-General of the United Nations would like answered include:

1. Are there particular countries(s) that are highly unusual in terms of the measured characteristics? If so, identify them.
2. Are there notable associations/relationships between some of the variables? (if so, describe them)
3. Is there a way to graphically represent the raw data for the 77 countries and draw conclusions about the data set from such a graph?
4. Can we find a few indices that describe the variation in the data set using a lesser dimension than the original set of variables? If so, what are those indices? Is there a convenient interpretation of any of the indices?
5. Can we graphically display the data in a low number of dimensions using such indices? What conclusions about the countries (individual countries or groups of countries) can you draw from such a graph?
6. Are there any countries that are similar or different from each other in any aspects that are surprising to you? What useful information, e.g., for a public relations campaign, could be gleaned from this data set as related to this?
7. What are any other potentially interesting aspects of the data set that may be gleaned for these data?

You will type a roughly 3-page report detailing your analysis of the data and your conclusions. Keep in mind that the report should be written for two audiences: the publications staff member of the secretary-general, who knows about world politics but is not an expert in statistics; and your supervisor, the head statistician, who will be judging you and deciding on your possible promotion based on the statistical competence of the report. Your report should be understandable and meaningful to both audiences.

You may include graphs that illustrate and/or support your findings. (The graphs do not have to count as part of the roughly 3-page length.) Do NOT include computer code within the main body of your report. This will be incomprehensible to the publications staff member and would only annoy her. You may include such code in an appendix if you wish.

The data for this problem are given at the link “Countries Data with names” on the course web page. At the top of the next page is some R code that may be helpful in reading in and managing the data:
2. You are working as a consulting statistician for an NBA basketball team that is preparing for roster planning and contract negotiations with several of its players. Data have been gathered on 176 players. The variables measured include several performance attributes of the players (data are actually from the 2011 season). The 13 variables measured on each player are: Age (player’s age in years), Starts (number of games the player started for the season), MinPerGame (average number of minutes the player played per game), FGPct (percentage of field goals attempted that the player successfully made – this includes close shots and “long-distance” shots), FG3Pct, (percentage of 3-point field goals attempted that the player successfully made – these are only “long-distance” shots), FTPct (percentage of free throws attempted that the player successfully made – these are unguarded medium-distance shots following an opponent’s foul), ReboundsPerMin (number of rebounds per minute; typically tall players have more chance to get rebounds), AssistsPerMin (number of assists per minute; typically quick ball-handlers have more chance to get assists), StealsPerMin (number of steals per minute; typically quick players have more chance to get steals), BlocksPerMin (number of blocked shots per minute; typically tall players have more chance to get blocks), TurnoversPerMin (number of turnovers per minute; typically players who often handle the ball have more chance to get turnovers), FoulsPerMin (number of fouls per minute), PointsPerMin (number of points scored per minute).

The questions that the team would like answered include:

1. Are there individual players who are highly unusual (in any way) based on the measured performance-related variables (those variables other than age, starts, and minutes per game)? If so, identify their names.
2. Are there notable associations/relationships between some of the variables? (if so, describe them)
3. Is there a way to graphically represent the raw data for the 176 players and draw conclusions about the data set from such a graph?
4. Are there a small number of underlying characteristics of players that the observed variables might be connected to? If so, determine how many latent characteristics there seem to be in this set of variables. Also, try to interpret them the best you can, with the aid of statistical techniques.
5. Can we graphically display the data in a low number of dimensions using such latent traits? What conclusions about the players (individual players or groups of players) can you draw from such graph(s)?
6. What are any other potentially interesting aspects of the data set?

You will type a roughly 3-page report detailing your analysis of the data and your conclusions. Keep in mind that the report should be written for two audiences: the basketball team’s client, who has a sense for numbers but is not an expert in statistics; and your own supervisor at the statistical consulting company, who will be judging you and deciding on your possible promotion based on the statistical competency of the report. Your report should be understandable and meaningful to both audiences.

You may include graphs that illustrate and/or support your findings. (The graphs do not have to count as part of the roughly 3-page length.) Do NOT include computer code within the main body of your report. This will be incomprehensible to the client and would only annoy him. You may include such code in an appendix if you wish.
The data for this problem are given at the link “NBA Players Data with names” on the course web page. Here is some R code that may be helpful in reading in and managing the data:

```r
NBA.full <- read.csv("http://people.stat.sc.edu/Hitchcock/NBAplayers530exam.txt", header=T)

attach(NBA.full)

NBA.names <- as.character(NBA.full[,1])

NBA.names.abb <- abbreviate(NBA.names)

NBA.data.numeric <- NBA.full[, -1]
```

**Grading Scale:**

Each problem will be worth 30 points, for a total of 60 points. For each problem, your report will be graded based on Writing, Analysis, and Context. For example:

**Writing** (out of 10 points): How organized, clearly written, comprehensible, and grammatically correct is the report? Would the client reading this report be confident that it was written by an educated, well-trained statistical scientist?

**Analysis** (out of 10 points): Were the graphs and data analyses appropriate for the problem? Were the analyses carried out correctly? Were your statistical conclusions about the data set sensible and clearly justified by numerical or graphical evidence?

**Context** (out of 10 points): Were the questions answered in terms of the variables of the data set? Although you are not an expert in the field as your client is, have you attempted to frame your conclusions and interpretations in a subject-matter context rather than treating the data as simply a meaningless set of numbers?