Your project should be submitted in a clean, typed, report-like format with a coversheet.

The goal of this project is to familiarize you with StatCrunch and examine different samples of data. Additionally, we will be using some of the samples you find in the second project. For this project, everyone may work alone, or work with the partner of your choice (1 only please). Everyone will be using the same data set. The due date is Friday, February 15th. All reports should be word-processed and stapled.

The grading rubric used for this project is as follows:

1. 75% for correct statistical techniques and corresponding conclusions.
2. 25% for creativity and professionalism – I’d like to see some attempt at a crisp, clearly written, well-organized and professional looking document. All graphics and tables should be labeled well and easy to find.

If you wish to register for StatCrunch, you can by completing the following steps:

2. Under “Student Options” choose the option “Purchase 6 months of access for $13.20”. Then complete and submit the registration form.
3. Do a search for the group “CSU STAT 307: Spring 2013: Section 001” and select join.

For a short (14 min) tutorial video that you may find instructive you can visit:
http://www.youtube.com/user/StatisticsForAll
The Setting:
In a study of oral home care practice, suppose that a certain researcher studies a census of size 2500 taken from a certain county. The oral hygiene status of all subjects in the county was examined using a plaque index with a range of 0 to 3, where 0 indicates that the subject has no soft plaque deposits, and 3 indicates that the subject has many soft plaque deposits. These responses are recorded in the data set file, “plaque index” found on RamCT, as well as on StatCrunch.

The Rationale:
You are going to play the role of the “typical” investigator who wishes to know the average plaque index of all people in this county, but does not have a full census of responses available. The project consists of 3 stages.

Stage 1 is the sampling stage: You will select 9 simple random samples from the target population, 3 each of sample sizes 10, 30, and 100.

Stage 2 is the estimation stage: You and your partner will calculate relevant descriptive statistics for each of the samples.

Stage 3: You and your partner will then obtain the true average plaque index by getting descriptive statistics for the entire population.

A) Please be sure to include a coversheet with your names (alphabetically by last name) in the upper right hand corner and a title somewhere appropriate.

B) Provide your written responses (see page 4) in the leading pages of your project document. The last pages of the project should contain all figures, output and tables. Please be sure that any figures, output, or tables are labeled appropriately with an identifying number and informative caption. At the very minimum you’ll need:

1) Relevant descriptive statistics for each sample. Don’t report useless statistics.

2) Descriptive stats for the entire target population (only 1 table required.)

3) (Multiple) Boxplots - for results of each sample as well as for the entire population. It’s probably a good idea to include several different multiple boxplots to help back up your conclusions.
Instructions for StatCrunch:

Open StatCrunch and log on. Select the “plaque index” data set

Part 1 - The sampling
A) You will draw 3 SRS of 10 responses from the population of plaque indices:
   1) From the menu bar select: Data > Sample Cols
      a) Sample size = 10
      b) Number of samples = 3
      c) Select “split across columns”
      d) Select “use preset generator”
      e) All other check boxes and radio buttons are left blank
      f) Click “Sample Column(s)”
      g) Rename the new column with something informative (but short)
   2) Repeat using sample sizes of 30 and 100.

Part 2 - Descriptive statistics
A) Obtain the descriptive statistics for each sample
   1) From the menu bar select: Stat > Summary Stats > Columns
      a) Select the 9 columns containing the sample results you got in Part 1. Leave the
         defaults as they are.
      b) Click “next”
      c) Deselect the “Variance” and “Range” - these are redundant when you have the
         standard deviation and the 5-number summary
      d) Click “Calculate”

Part 3 - Descriptive statistics for the census
A) Obtain the descriptive statistics for the census
   1) From the menu bar select: Stat > Summary Stats > Columns
      a) Select the column containing the census results (“plaque index”). Leave the
         defaults as they are.
      b) Click “next”
      c) Deselect the “Variance”, “Range”, and “Std. Dev.”
      d) Select “Unadj. Std. Dev.”
      e) Click “Calculate”

Part 4 - Boxplots
A) Obtain boxplots and multiple boxplots
   1) From the menu bar select: Graphics > Boxplot
      a) Select the columns containing the data you wish to be included in your (multiple)
         boxplot.
      b) Click “next”
      c) Select “Use fences to identify outliers”
      d) Click “Calculate”

NOTE: You may want to save your data in StatCrunch, as we will be using it again in Project 2.
Disregard the fact, for the moment, that we have a complete census of data and put your self in the role of an investigator who wants to estimate the true average plaque index for all people in the county. You have a complete list of everyone available - statisticians call this kind of list a sampling frame - and you are going to select a SRS from this list. Answer the following questions. Please note that your answers should be written in clear, complete sentences.

1) What is the parameter of interest?
2) What is the variable of interest?
3) Describe the variable type and level.
4) Define the target population.
5) Define the sampled population.
6) Are the target and sampled population the same?
7) Consider each group of 3 samples of the same sample size. Why do you think the summary statistics of the three samples within each of the groups differ? Do you think one sample is better than the others?
8) Consider each group of 3 samples of the same sample size. What differences between the three groups, if any, do you notice?
9) Of the three different sample sizes used, which do you think is the best for estimating the true (census) average plaque index? Why?