The Net Promoter Score (NPS) survey asks users how likely they would be to recommend our products to a friend or colleague, using a scale from 0 (not at all likely) to 10 (definitely). Indeed sends this survey to its employers who post jobs on the website. In this exercise, we're going to ask you to manipulate and analyze data about employers using Indeed and their NPS survey responses.

Please use R or Python (any packages/functions you are comfortable with) to answer the questions. There are four datasets (.csv) available to you. A description of the datasets is below:

- *nps,* where a row represents a survey response and columns include:
    - **advertiser_id** is an identifier for a company
    - **nps_score** is the rating given
- *country,* where a row represents a company and columns include:
    - **advertiser_id** is an identifier for a company,
    - **billing_country_code** is the region in which a company is located
    - **nps_score** is the rating given
- *jobs,* where a row represents a company and columns include:
    - **metadataAdvertiserID** is an identifier for a company, and is identical to advertiser_id but has a slightly different name ,
    - **distinct(jobid)** is the number of jobs each company has posted on Indeed
    - **visibilitylevel** is a flag for how visible a job on Indeed is to job seekers - some jobs are eligible for more visibility than others
- *verbatims*, which includes open-ended text responses to a form field in our survey

The questions are designed to become progressively more difficult, but each solution is achievable using common analysis packages.

Deliverables to the Indeed team

1. **A one-page .pdf with written answers:** Please answer the questions below in a text document (e.g. .doc, .docx, Google Doc), and save them in a .pdf format. All of your answers must together fit on one page with standard font sizing (10-12) and margins (~1"). Your answers should use plain language since the reviewer may not have technical statistics knowledge.
2. **A source code file (.R, .Rmd, or .py):** You will need to submit your source code file (.R, .Rmd, or .py format) that you used to arrive at your answers. Use comments to explain your thinking behind the choice of function/coding method. The reviewer will have technical statistics and coding knowledge, but make it as clear as possible how you are thinking through your answer.

Data Exercise questions

1. Determine if there are differences in NPS score by country.
    a. Are any differences statistically significant, and how did you decide?
2. Is there a difference in NPS by number of jobs posted?
    a. Are any differences statistically significant, and how did you decide?
3. Company leadership is interested in segmenting NPS data by company size, but we don't collect the size of the company. What segmentation strategy would you suggest based on the data available?
    a. Why did you choose that segmentation strategy?
    b. What would you suspect are 1-2 dangers from this approach, and how would you convey these to company leadership? Why is your strategy still viable despite those dangers?
4. What strategies/methodologies would you use to extract the most common themes in verbatim responses from customers in the NPS survey?
    a. What are some themes in these responses? If you're able to execute some code to find some themes, great, but don't spend too much time here. It's more important to demonstrate your understanding of methodologies.