Assignment 4: Static Visualization

In this assignment, you will design three visualizations for a dataset about the human body and movement, and provide a rigorous rationale for your design choices. You should in theory be ready to explain the contribution of every pixel in the display. You are free to use any graphics or charting tool you please – including drafting it by hand. However, you may find it most instructive to create the charts from scratch using a graphics API of your choice.

Datasets

Choose **one** of the two datasets below. Both are large, so you will need to choose which variables and which people your questions are about. Every variable is described in that dataset's codebook.

Dataset A: NHANES 2013–2014 Movement, Fitness and Health

Every two years, the U.S. National Health and Nutrition Examination Survey (NHANES) examines a national sample of people of all ages, informing everything from physical activity guidelines to growth charts and clinical reference values. Each participant is interviewed at home, examined in a mobile exam center, has blood drawn, and wears a wrist accelerometer day and night for a week. This dataset brings those measures together for the 2013–2014 cycle. The data is a CSV (comma-separated values) file with one row per person, describing demographics (age, sex, race and ethnicity, education, income), body size, resting heart rate and blood pressure, grip strength, DXA body composition (fat, lean and bone), self-reported work, transport and leisure activity, sitting and screen time, sleep, general health, blood markers (HbA1c, cholesterol), and accelerometer movement, including total movement per day, weekday vs. weekend movement, and movement in each hour of the day. There are 10,175 people and 84 variables. Not every measure was collected for every age group, so many cells are blank by design.

Dataset: nhanes_2013_2014_kinesiology.csv

Codebook: codebook_nhanes.md

Source: U.S. Centers for Disease Control and Prevention, National Center for Health Statistics (NHANES), https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2013

Dataset B: 1000 Genomes Exercise and Physiology Gene Variants

The 1000 Genomes Project sequenced the whole genomes of people from 26 populations on five continents, creating a widely used open catalog of how human DNA varies around the world. This dataset extracts 18 single-letter DNA variants in genes studied in exercise science and physiology: muscle fiber function (*ACTN3*, the "sprinter gene"), mitochondria and endurance (*PPARGC1A*), muscle energy (*AMPD1*), oxygen sensing (*HIF1A*), lactate transport (*SLC16A1*), tendons and ligaments (*COL5A1*), blood pressure (*AGT*), adrenaline response (*ADRB2*), antioxidant defense (*SOD2*), vitamin D and bone (*VDR*, *SLC24A5*), motor learning and the brain (*BDNF*, *COMT*, *APOE*), obesity (*FTO*), and diet (*LCT*, lactase; *CYP1A2*, caffeine). It comes as two CSV files. The individual file describes each person in terms of sex, population, continental group, their two-letter genotype at each variant, and how many copies they carry of the allele of interest (0, 1 or 2). There are 2,504 people. The frequency file summarizes the same variants as allele frequencies for each of the 26 populations and five continental groups, for a total of 1,093 rows.

Dataset: genomes_1000g_individuals.csv, genomes_1000g_population_frequencies.csv

Codebook: codebook_genomes.md

Source: 1000 Genomes Project (phase 3), https://www.internationalgenome.org/, via the Ensembl REST API, https://rest.ensembl.org/

Your Tasks

- Choose three related questions that you would like to answer through data visualization. Together, the questions should explore a broader topic in kinesiology, such as movement, exercise, health, or human performance, in an individual, clinical, or population context. Use the same dataset to address all three questions.
- For each question, design a static visualization (i.e., a single image) that effectively communicates an answer. Use the question as the title of the visualization. You will submit three images in total.
- Write one paragraph for each visualization explaining your design choices and how the visualization helps answer its question.
- Write one concluding paragraph that synthesizes what the three visualizations reveal about the broader topic.

Your submission should include three visualization images and four paragraphs in total.

While you must use one of the datasets given, you are free to transform the data as you see fit. Such transforms may include (but are not limited to) log transformation, computing percentages or averages, grouping elements into new categories (such as age groups or BMI categories), or removing unnecessary variables or records. You are also free to incorporate external data as you see fit. Each chart image should be interpretable without recourse to your paragraphs. Do not forget to include title, axis labels or legends as needed!

In each visualization paragraph, provide a rigorous rationale for your design decisions. State what aspect of the data you are trying to communicate most effectively, and which aspects might be obscured or down-played by your design. Justify the visual encodings you chose and why they suit the data and your question, including the choice of visualization type, size, color, scale, and other visual elements, as well as the use of sorting or other data transformations. How do these decisions facilitate effective communication?

In your concluding paragraph, explain how the three questions relate, what the three visualizations show together that no single one shows alone, and what remains unanswered.

Grading

The assignment score is out of a maximum of 5 points. Each of the five rubric components below is worth 1 point and is judged across all three visualizations and your four paragraphs. We will determine scores by judging both the soundness of your design and the quality of the write-up. We will also look for consideration of audience, message and intended task.

We will use the following rubric to grade your assignment. Note, rubric cells may not map exactly to specific point scores.

Component	Excellent	Satisfactory	Poor
Data Question	An interesting question (i.e., one without an immediately obvious answer) is posed. The visualization provides a clear answer.	A reasonable question is posed, but it is unclear whether the visualization provides an answer to it.	Missing or unclear question posed of the data.
Mark, Encoding, and Data Transforms	All design choices are effective. The visualization can be read and understood effortlessly.	Design choices are largely effective, but minor errors hinder comprehension.	Ineffective mark, encoding, or data transformation choices are distracting or potentially misleading.
Titles & Labels	Titles and labels helpfully describe and contextualize the visualization.	Most necessary titles and labels are present, but they could provide more context.	Many titles or labels are missing, or do not provide human-understandable information.
Design Rationale	Well crafted write-up provides reasoned justification for all design choices.	Most design decisions are described, but rationale could be explained at a greater level of detail.	Missing or incomplete. Several design choices are left unexplained.
Creativity & Originality	You exceeded the parameters of the assignment, with original insights or a particularly engaging design.	You met all the parameters of the assignment.	You met most of the parameters of the assignment.

Submission Details

This is an individual assignment. You may not work in groups. Your completed assignment is due on [day] [date], by [time].

Submit your assignment using [the submission form]. The form expects your three visualizations to be three separate images (each either a .png or .jpg), along with your four paragraphs. Please make sure each image is sized for a reasonable viewing experience -- readers should not have to zoom or scroll in order to effectively view your submission!

Resubmissions. Resubmissions will be regraded by the instructor, and you may earn back up to 50% of the points lost in the original submission. To resubmit this assignment, please use [the resubmission form] and follow the same submission process described above. Include a short 1 paragraph description summarizing the changes from the initial submission. Resubmissions without this summary will not be regraded. Resubmissions will be due by [time] on [day], [date]. Slack days may not be applied to extend the resubmission deadline. The teaching staff will only begin to regrade assignments once the Final Project phase begins, so please be patient.