SOCI 269

Coding Assignment

Sakeef M. Karim

*Amherst College*

## Basic Expectations

As noted in your [syllabus](https://soci269.netlify.app/), you are required to submit a short coding assignment by Monday, March 10th at 8:00 PM. For this assignment, you will clean a [dataset](#thedata) in , report basic descriptive statistics, and create simple data visualizations. You must also include your script file (i.e., a .R document) as part of your submission. Once you’re done, please [submit your materials via Moodle.](https://moodle.amherst.edu/mod/assign/view.php?id=934102)

|  |
| --- |
| You Must Submit Two Separate Files |
| Please remember to submit (i) the code you used to complete the assignment **along with** (ii) a text-based summary of your results. |

## The Data

### Description

You will be working with a truncated version of the 2010 *General Social Survey* (henceforth, GSS). The dataset was prepared using the [{gssr}](https://kjhealy.github.io/gssr/) package in .

You can access the data through one of three channels:

1. By copying and pasting the script below directly into RStudio:

readRDS(url("https://github.com/sakeefkarim/intro\_quantitative\_sociology/raw/refs/heads/main/data/assignments/coding%20assignment/gss\_2010\_truncated.rds"))

1. By [downloading](./data/gss_2010_truncated.rds)  the .rds file.
2. By cloning our companion [GitHub repository](https://github.com/sakeefkarim/intro_quantitative_sociology/tree/main).

### Variables

Learn more about the variables in your data by using the interactive table [embedded online](https://soci269-a1.netlify.app/#variables). This table includes data on all variables with labels in the broader (i.e., non-coarsened) 2010 GSS.

## Your Tasks

1. Report the mean for all numeric variables in the data—*with* and *without* weights.[[1]](#footnote-1)

|  |
| --- |
| **Hint**  You may want to explore the [weighted.mean()](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/weighted.mean) function. |

1. Report the median age of all respondents *by* race and sex. Concretely, your estimates should provide the median age of Black women, “Other” men *etc.* These results do **not** have to be weighted. That said, if you *want* to generate weighted medians, feel free to explore the [Hmisc::wtd.quantile()](https://rdrr.io/cran/Hmisc/man/wtd.stats.html) function.
2. Report the share of respondents who are Democrats—including Independents who *lean* Democrat and those who do not consider themselves “strong” Democrats. Once again, these results do **not** have to be weighted.[[2]](#footnote-2)
3. Explore the hrsrelax, mntlhlth and physhlth variables. What do they refer to? Are they meaningfully patterned by age, race, religion, sex, sexuality and their many intersections? Using ggplot2, generate **two** simple visualizations that provide preliminary insights based on your exploratory assessments and hunches.

|  |
| --- |
| **Hint**  You may want to use [facet\_wrap()](https://ggplot2.tidyverse.org/reference/facet_wrap.html) or [facet\_grid()](https://ggplot2.tidyverse.org/reference/facet_grid.html) to simplify your story. |

1. What does the letin1a variable capture? Generate a *third* visualization using ggplot2 that illustrates how letin1a may be socially patterned.

## Formatting Guidelines

You are free to prepare the exposition for your assignment in Microsoft Word, Google Docs, LaTeX, RMarkdown or Quarto. Concretely, this means you can submit your text-based summary as a .docx file or as a . Please use complete sentences to proffer your basic arguments and interpret the results you present. To facilitate interpretation, generate simple tables or plots to summarize descriptive results (i.e., Questions [1-3](#tasks).)

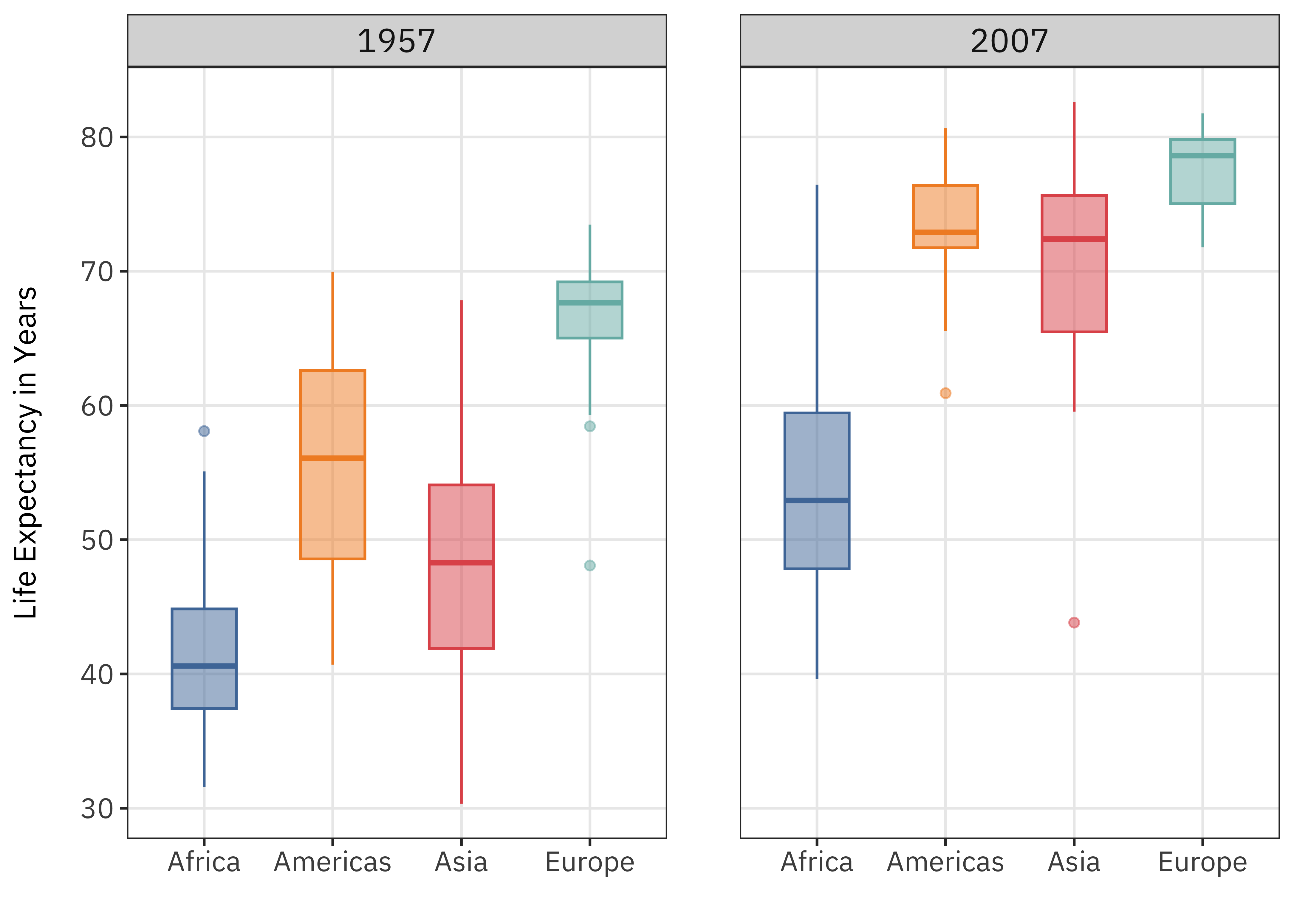
If you decide to include references, please use [APA](https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/general_format.html) *or* [ASA](https://owl.purdue.edu/owl/research_and_citation/asa_style/index.html) citation styles to manage references and bibliographies.[[3]](#footnote-3) More generally, *you must use subheadings to organize your arguments*.

## 

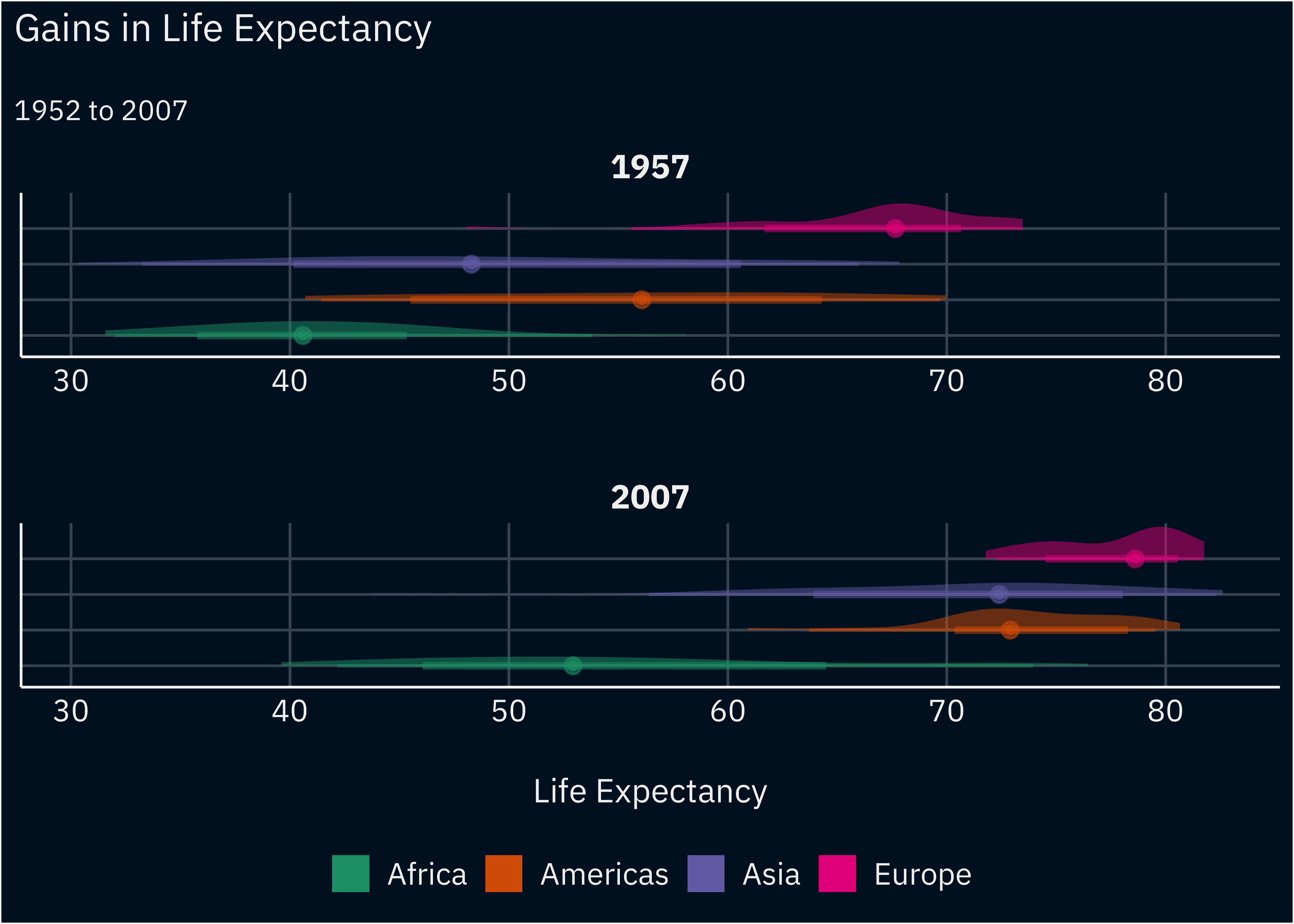
## Bonus

|  |
| --- |
| These Are Bonus Questions |
| You do **not** have to submit answers to these questions. They are *bonus* items for students with prior exposure to ggplot2 (or those who want more practice). |

1. Reproduce the plot below using gapminder and the ggthemes package.



1. Reproduce the plot below using the see package and geoms from ggdist.



1. Produce a map of Greater Boston that speaks to the racial diversity of the city.
2. Use dplyr functions to merge the coarsened 2010 GSS (gss\_2010\_truncated.rds) with interesting variables from the broader 2010 GSS file (gss\_2010.rds).

1. You do not need to provide means for the weighting variable. [↑](#footnote-ref-1)
2. You are, however, free to produce weighted estimates. [↑](#footnote-ref-2)
3. If you haven’t done so already, you may want to invest in [Zotero](https://www.zotero.org/) to manage your citations. [↑](#footnote-ref-3)