Project Proposal

Due November 17 at 11:59pm

Peter de Guzman, Sizhe Chen, Jennifer Li, Nakiyah Dhariwala

Load Packages

Dataset 1 (top choice)

Data source: The 2020 Collaborative Multi-racial Post-Election Survey (CMPS)

Brief description: The survey has an unweighted sample size of 17,545 respondents. Each row represents a unique respondent. The data is collected through an online survey with 800 unique questions. The survey's main focus is on attitudes about the 2020 Election and candidates, experiences with racism, policy attitudes, immigration, and personal experiences with civic engagement across many facets of American life.

The 2020 CMPS was offered in English, Spanish, Chinese (simplified), Chinese (traditional), Korean, Vietnamese, Arabic, Urdu, Farsi and Haitian Creole. The survey dataset includes Black, White, Latino and Asian respondents as well as additional oversamples of respondents from hard-to-reach populations including, Afro-Latinos, Black immigrants, Native Americans, Native Hawaiians, Muslims and people who identify as LGBTQ.

Research question 1: Are Americans who believe their race is important to their identity more likely to be interested in politics than those that are not?

- Outcome variable (include the name/description and type of variable): The outcome variable is an ordinal variable labeled Q29. The survey question is "Some people are very interested in politics while other people can't stand politics how about you? are you…". The response options are in a five-point scale from "Not at all interested in politics" to "Very interested in politics."

The main predictor variable is "Q271" which asks "How important is being [RACIAL GROUP] to your identity?". The response options are a five-point scale from "Not at all important" to "Extremely important". Other potential predictors include survey questions on respondent's self-perception of race and discrimination, and their own attitudes towards other racial groups. In addition, there are multiple variables that could be used as demographic controls such

as age, gender, education, income, and party identification. The variables not shown in the output include responses to the rest of the 800 survey questions, covering topics such as policy attitudes, civic engagement, and political participation.

Variable Type: Ordinal

Research question 2: Controlling for other demographic variables such as age and gender, how do experiences with racism influence the importance respondents place on policy issues related to racial and ethnic discrimination?

- Outcome variable (include the name/description and type of variable): Stopping discrimination against racial/ethnic minorities (Q1R5) — a binary variable indicating whether respondents view it as an important issue.

Variable Type: Binary.

Interaction Term Idea: Include an interaction term between race/ethnicity (S2RACER) and experiences with racism to see if the influence of racism experiences differs across racial/ethnic groups.

Load the data and provide a `glimpse()`:

```
'data.frame':   50 obs. of  1490 variables:
 $ RECORD           : num  51853 13765 13630 59404 25918 ...
 $ UUID             : Factor w/ 17545 levels "002krf5kb6aywfam",..: 1 2 3 4 5 6 7 8 9 10 ..
 $ DATE             : Factor w/ 13110 levels "01-AUG-2021 00:33:00",..: 10144 8645 8653 137
 $ Q29              : Factor w/ 4 levels "(1) Very interested in politics",..: 3 3 2 1 1 3 2
 $ Q271             : Factor w/ 5 levels "(1) Extremely important",..: 3 2 2 2 2 3 1 3 1 1
 $ WEIGHT           : num  1.401 1.954 0.197 0.454 0.812 ...
 $ AGE1_HID         : Factor w/ 2 levels "(1) 43","(2) 45": 1 2 1 2 2 2 2 2 2 1 ...
 $ AGE2_HID         : Factor w/ 2 levels "(1) 43","(2) 45": 2 1 2 1 1 1 1 1 1 2 ...
 $ S1               : Factor w/ 10 levels "(01) English",..: 1 1 1 1 1 1 1 2 1 1 ...
 $ S2_RACER1        : Factor w/ 2 levels "(0) NO TO: White",..: 1 2 1 1 1 1 2 1 1 2 ...
 $ S2_RACER2        : Factor w/ 2 levels "(0) NO TO: Hispanic or Latino",..: 1 1 2 2 1 1 1 2
 $ S2_RACER3        : Factor w/ 2 levels "(0) NO TO: Black or African American",..: 2 1 1 1
 $ S2_RACER4        : Factor w/ 2 levels "(0) NO TO: Asian American",..: 1 1 1 1 1 2 1 1 2 1
 $ S2_RACER5        : Factor w/ 2 levels "(0) NO TO: American Indian/Native American",..: 1
 $ S2_RACER6        : Factor w/ 2 levels "(0) NO TO: Arab, Middle Eastern or North African"
 $ S2_RACER7        : Factor w/ 2 levels "(0) NO TO: Native Hawaiian",..: 1 1 1 1 1 1 1 1 1
 $ S2_RACER8        : Factor w/ 2 levels "(0) NO TO: Not Hawaiian, but other Pacific Islande
 $ S2_RACE_PRIME    : Factor w/ 8 levels "(1) White","(2) Hispanic or Latino",..: 3 1 2 2 3
 $ MAIN_ETH_QUOTA   : Factor w/ 7 levels "(1) White","(2) Latino",..: 3 1 2 5 3 4 3 2 4 2 .
 $ S2_HISPANICR1    : Factor w/ 2 levels "(0) NO TO: No, not Hispanic or Latino",..: 2 2 NA
 $ S2_HISPANICR2    : Factor w/ 2 levels "(0) NO TO: Yes, myself",..: 1 1 NA NA 1 1 1 NA 1 N
 $ S2_HISPANICR3    : Factor w/ 2 levels "(0) NO TO: Yes, one of my parents",..: 1 1 NA NA 1
```

```
 $ S2_HISPANICR4     : Factor w/ 2 levels "(0) NO TO: Yes, one of my grandparents",..: 1 1 N
 $ S2_HISPANICR5     : Factor w/ 2 levels "(0) NO TO: Yes, another more distant relative",..
 $ S2_NATIVER1       : Factor w/ 2 levels "(0) NO TO: No, not Native American",..: 2 2 2 1 2
 $ S2_NATIVER2       : Factor w/ 2 levels "(0) NO TO: Yes, myself",..: 1 1 1 1 1 1 1 1 1 1 .
 $ S2_NATIVER3       : Factor w/ 2 levels "(0) NO TO: Yes, one of my parents",..: 1 1 1 2 1 1
 $ S2_NATIVER4       : Factor w/ 2 levels "(0) NO TO: Yes, one of my grandparents",..: 1 1 1
 $ S2_NATIVER5       : Factor w/ 2 levels "(0) NO TO: Yes, another more distant relative",..
 $ S2_AFROLAT_TEXT   : Factor w/ 2 levels "(1) If you do not have any ancestry from Latin Ame
 $ S2_AFROLATR1      : Factor w/ 2 levels "(0) NO TO: No, not Afro-Latino",..: 2 NA 2 2 2 NA
 $ S2_AFROLATR2      : Factor w/ 2 levels "(0) NO TO: Yes, myself",..: 1 NA 1 1 1 NA 1 1 NA 1
 $ S2_AFROLATR3      : Factor w/ 2 levels "(0) NO TO: Yes, one of my parents",..: 1 NA 1 1 1
 $ S2_AFROLATR4      : Factor w/ 2 levels "(0) NO TO: Yes, one of my grandparents",..: 1 NA 1
 $ S2_AFROLATR5      : Factor w/ 2 levels "(0) NO TO: Yes, another more distant relative",..
 $ S3                : Factor w/ 6 levels "(01) Straight, that is, not gay or lesbian",..: 3
 $ S3R40E            : Factor w/ 2 levels "                    ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ S3B               : Factor w/ 4 levels "(1) Man","(2) Woman",..: 2 2 2 1 2 1 1 1 2 2 ...
 $ S3BR40E           : Factor w/ 2 levels "                    ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ S3C               : Factor w/ 3 levels "(01) Yes","(02) No",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ S4                : Factor w/ 51 levels "(01) Alaska",..: 15 32 10 33 11 11 33 10 5 6 ...
 $ S5                : num  2000 1964 1953 1997 1949 ...
 $ S5_AGE            : Factor w/ 7 levels "(1) 16-17","(2) 18-29",..: 2 5 6 2 7 2 2 7 4 2 ..
 $ S6                : Factor w/ 3 levels "(1) Yes, registered to vote",..: 2 1 1 2 1 3 2 1 2
 $ S6_NON            : Factor w/ 3 levels "(1) Correct, I am not currently registered to vote
 $ S7                : Factor w/ 3 levels "(1) United States",..: 1 1 1 1 1 2 1 2 2 1 ...
 $ S8                : Factor w/ 16 levels "(01) Chinese",..: NA NA NA NA NA 5 NA NA 3 NA ..
 $ S8R160E           : Factor w/ 2 levels "                    ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ ASIAN_ETH         : Factor w/ 8 levels "(1) Chinese",..: NA NA NA NA NA 5 NA NA 3 NA ...
 $ S8_PACIFIC        : Factor w/ 8 levels "(1) Native Hawaiian",..: NA NA NA NA NA NA NA NA 1
 $ S8_PACIFICR80E    : Factor w/ 2 levels "                    ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ S9                : Factor w/ 19 levels "(01) Roots in America for many generations",..: 1
 $ S9R190E           : Factor w/ 2 levels "                    ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ S10               : Factor w/ 22 levels "(01) Argentina",..: NA NA 21 10 NA NA NA 6 NA 12
 $ S10R220E          : Factor w/ 2 levels "                    ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ S10_MEX           : Factor w/ 2 levels "(1) Yes, have Mexican ancestry",..: NA NA 2 NA NA
 $ S11               : Factor w/ 15 levels "(01) Navajo",..: NA NA NA 12 NA NA NA NA NA NA .
 $ S11R130E          : Factor w/ 2 levels "                    ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ S12               : Factor w/ 29 levels "(01) Afghanistan",..: NA NA NA NA NA NA NA NA NA
 $ S12R270E          : Factor w/ 2 levels "                    ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ S2_M              : Factor w/ 2 levels "(1) Yes","(2) No": 1 2 2 2 2 2 1 2 2 2 ...
 $ S13               : Factor w/ 7 levels "(1) Grades 1-8",..: 1 3 7 6 5 3 3 5 7 3 ...
 $ S14               : Factor w/ 5 levels "(1) Large urban area",..: 5 5 1 4 4 3 2 1 2 2 ...
 $ S15               : Factor w/ 1 level "MASKED BY ICPSR": 1 1 1 1 1 1 1 1 1 1 ...
 $ Q58R1             : Factor w/ 2 levels "(0) NO TO: Catholic",..: 1 1 1 2 1 1 1 2 1 2 ...
```

```
 $ Q58R2              : Factor w/ 2 levels "(0) NO TO: Protestant (such as Methodist, Presbyte
 $ Q58R3              : Factor w/ 2 levels "(0) NO TO: Other Christian",..: 1 1 2 1 1 1 1 1 1 1
 $ Q58R4              : Factor w/ 2 levels "(0) NO TO: Jewish",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Q58R5              : Factor w/ 2 levels "(0) NO TO: Muslim",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Q58R6              : Factor w/ 2 levels "(0) NO TO: Hindu",..: 1 1 1 1 1 1 1 1 1 2 1 ...
 $ Q58R7              : Factor w/ 2 levels "(0) NO TO: Buddhist",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Q58R8              : Factor w/ 2 levels "(0) NO TO: Mormon/LDS",..: 1 1 1 1 1 1 1 1 1 1 ..
 $ Q58R9              : Factor w/ 2 levels "(0) NO TO: Folk religion",..: 1 1 1 1 1 1 1 1 1 1 1
 $ Q58R10             : Factor w/ 2 levels "(0) NO TO: Daoist",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Q58R11             : Factor w/ 2 levels "(0) NO TO: Shamanism",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Q58R12             : Factor w/ 2 levels "(0) NO TO: Ancestral religion",..: 1 1 1 1 1 1 1 1
 $ Q58R13             : Factor w/ 2 levels "(0) NO TO: Spiritual but not religious",..: 1 1 1
 $ Q58R14             : Factor w/ 2 levels "(0) NO TO: Atheist",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Q58R15             : Factor w/ 2 levels "(0) NO TO: Agnostic",..: 1 1 1 1 1 2 1 1 1 1 ...
 $ Q58R16             : Factor w/ 2 levels "(0) NO TO: None",..: 2 1 1 1 1 1 1 1 1 1 ...
 $ Q58R17             : Factor w/ 2 levels "(0) NO TO: Other (SPECIFY)",..: 1 1 1 1 1 1 1 1 1 1
 $ Q58R17OE           : Factor w/ 2 levels "             ",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Q3B_PLURAL         : Factor w/ 4 levels "(1) men","(2) women",..: 2 2 2 1 2 1 1 1 2 2 ...
 $ Q3B_PLURALR4OE     : Factor w/ 1 level " ": 1 1 1 1 1 1 1 1 1 1 ...
 $ Q1R1               : Factor w/ 2 levels "(0) NO TO: Responding to Coronavirus / COVID-19",
 $ Q1R2               : Factor w/ 2 levels "(0) NO TO: Lowering the costs of health care",..:
 $ Q1R3               : Factor w/ 2 levels "(0) NO TO: Improving wages and incomes",..: 1 1 2
 $ Q1R4               : Factor w/ 2 levels "(0) NO TO: Unemployment / Creating more jobs",..:
 $ Q1R5               : Factor w/ 2 levels "(0) NO TO: Stopping discrimination against racial/
 $ Q1R6               : Factor w/ 2 levels "(0) NO TO: Criminal justice and police reform",..
 $ Q1R7               : Factor w/ 2 levels "(0) NO TO: Protecting immigrant rights",..: 1 1 1
 $ Q1R8               : Factor w/ 2 levels "(0) NO TO: Improving K-12 education and schools",
 $ Q1R9               : Factor w/ 2 levels "(0) NO TO: Lowering taxes",..: 1 1 1 1 1 1 1 2 1 1
 $ Q1R10              : Factor w/ 2 levels "(0) NO TO: Decrease government spending",..: 1 1 1
 $ Q1R11              : Factor w/ 2 levels "(0) NO TO: Combating climate change and pollution"
 $ Q1R12              : Factor w/ 2 levels "(0) NO TO: Creating more affordable housing",..: 1
 $ Q1R13              : Factor w/ 2 levels "(0) NO TO: Addressing mass shootings and gun safet
 $ Q1R14              : Factor w/ 2 levels "(0) NO TO: Lowering the cost of college",..: 1 1 1
 $ Q1R15              : Factor w/ 2 levels "(0) NO TO: Protecting women",..: 1 1 1 1 1 1 1 1 1
  [list output truncated]
 - attr(*, "variable.labels")= Named chr [1:1490] "record: Record number" "uuid: Participant
  ..- attr(*, "names")= chr [1:1490] "RECORD" "UUID" "DATE" "S1" ...
 - attr(*, "codepage")= int 28591
```
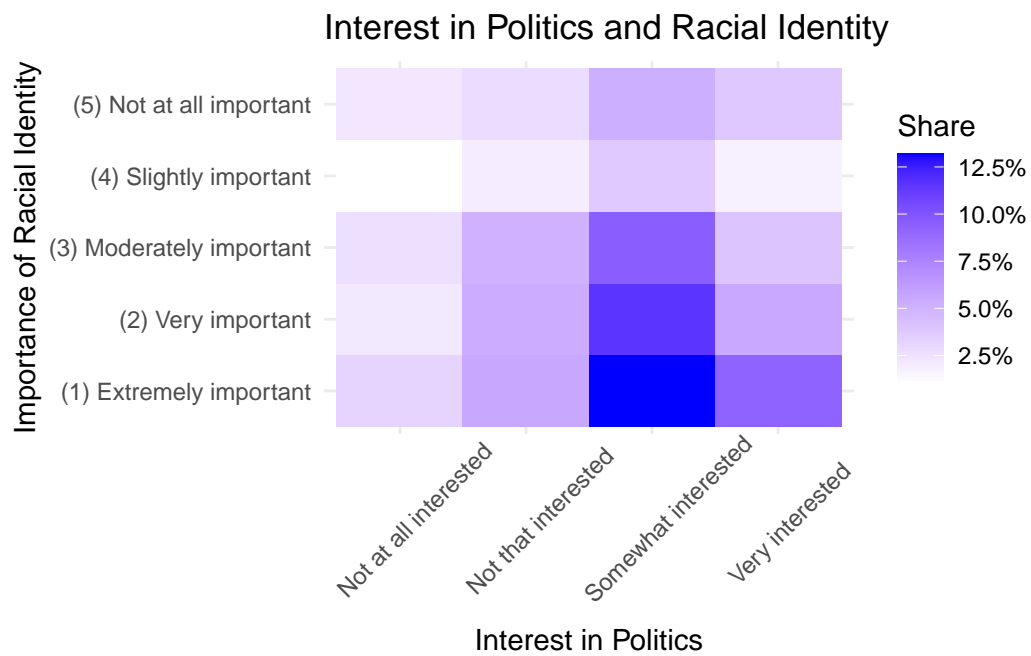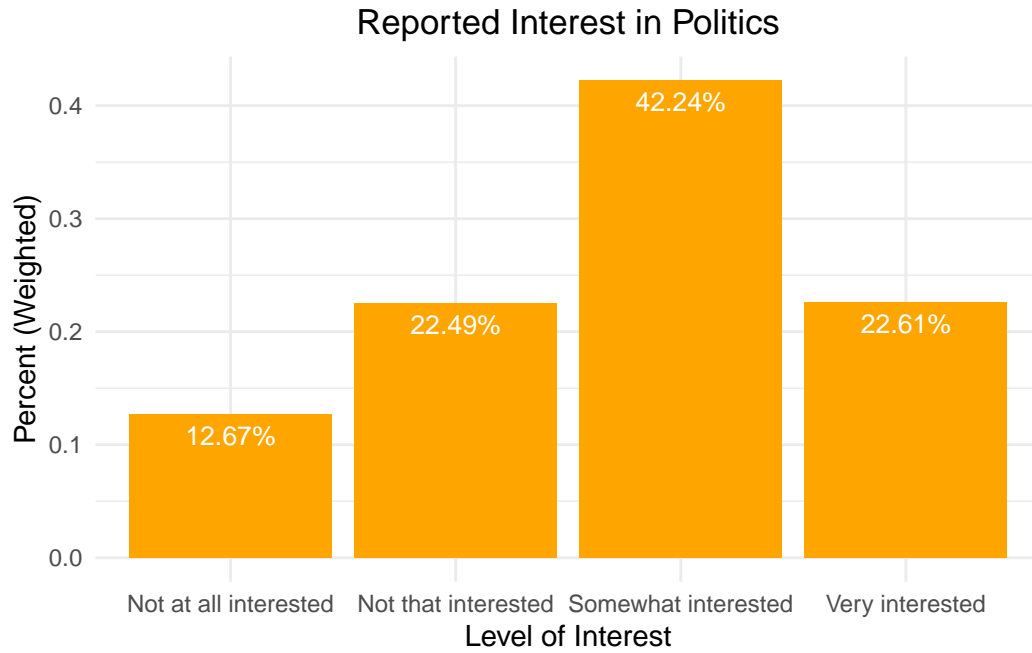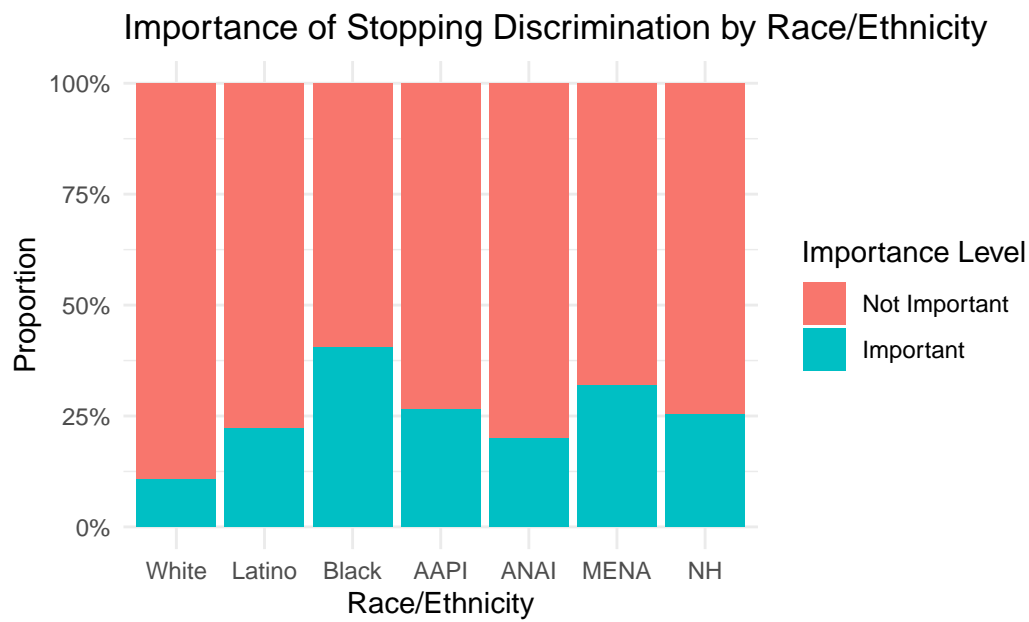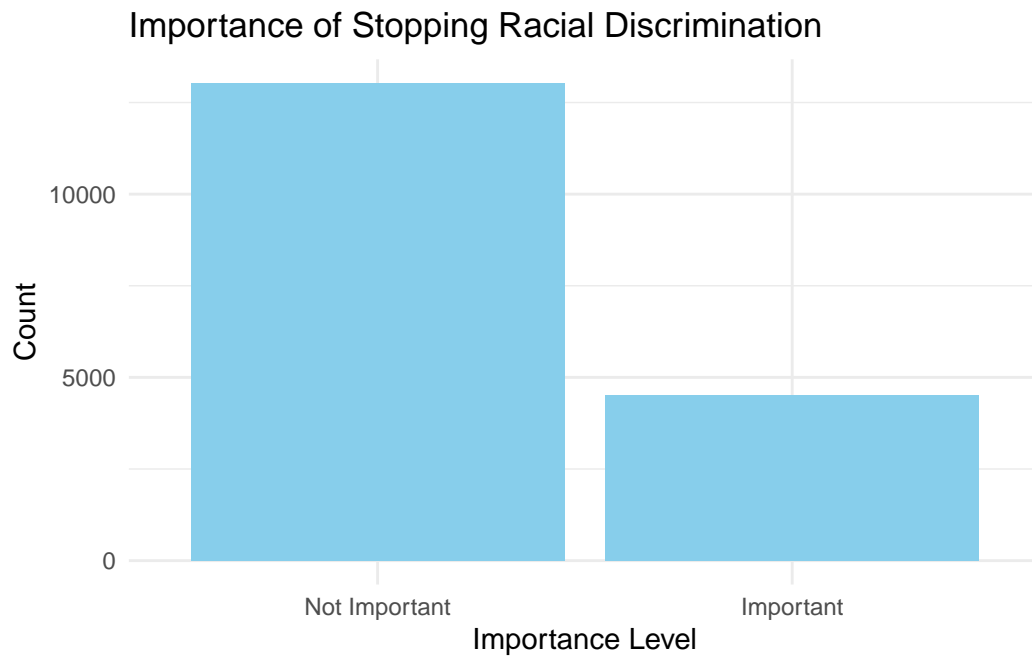
Exploratory Plots: Dataset 1: Research Question 1

Reported Interest in Politics

Interest in Politics and Racial Identity

Importance of Stopping Racial Discrimination

Importance of Stopping Discrimination by Race/Ethnicity

Asian American/Pacific Islander, MENA = Middle Eastern/North African

Dataset 2

Data source: Global Health Observatory (GHO) data repository under the World Health Organization (WHO), with additional economic data sourced from the United Nations website. https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

Brief description: There are 2938 rows and 22 columns, representing health and social/economic indicators for 193 countries over 15 years. Each row corresponds to a specific country-year combination.

Research question 1: How does the number of years a person spends in school influence whether a country is classified as developed or developing?

- Outcome variable (include the name/description and type of variable): "Status" –> Binary variable that categorizes countries into "Developed" or "Developing."

Research question 2: How do factors such as life expectancy, education, healthcare expenditure, and population affect GDP across developing countries?

- Outcome variable (include the name/description and type of variable): GDP: Numeric Variable
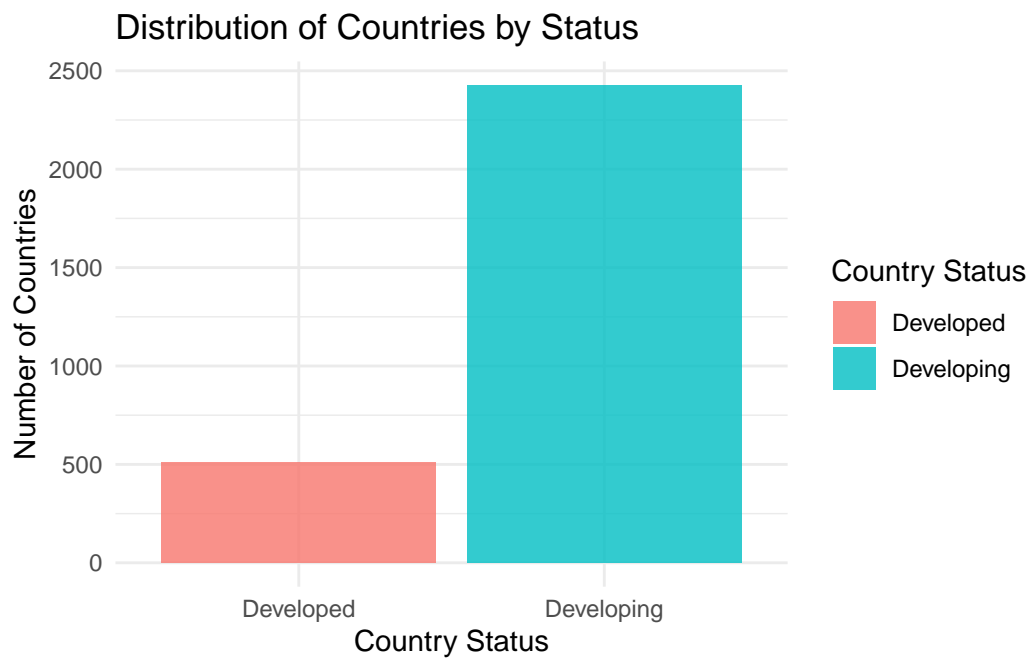
Load the data and provide a `glimpse()`:

```
Rows: 2,938
Columns: 22
$ Country                <chr> "Afghanistan", "Afghanistan", "Afghani~
$ Year                   <int> 2015, 2014, 2013, 2012, 2011, 2010, 20~
$ Status                 <chr> "Developing", "Developing", "Developin~
$ Life.expectancy        <dbl> 65.0, 59.9, 59.9, 59.5, 59.2, 58.8, 58~
$ Adult.Mortality        <int> 263, 271, 268, 272, 275, 279, 281, 287~
$ infant.deaths          <int> 62, 64, 66, 69, 71, 74, 77, 80, 82, 84~
$ Alcohol                <dbl> 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.~
$ percentage.expenditure <dbl> 71.279624, 73.523582, 73.219243, 78.18~
$ Hepatitis.B            <int> 65, 62, 64, 67, 68, 66, 63, 64, 63, 64~
$ Measles                <int> 1154, 492, 430, 2787, 3013, 1989, 2861~
$ BMI                    <dbl> 19.1, 18.6, 18.1, 17.6, 17.2, 16.7, 16~
$ under.five.deaths      <int> 83, 86, 89, 93, 97, 102, 106, 110, 113~
$ Polio                  <int> 6, 58, 62, 67, 68, 66, 63, 64, 63, 58,~
$ Total.expenditure      <dbl> 8.16, 8.18, 8.13, 8.52, 7.87, 9.20, 9.~
$ Diphtheria             <int> 65, 62, 64, 67, 68, 66, 63, 64, 63, 58~
$ HIV.AIDS               <dbl> 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1~
$ GDP                    <dbl> 584.25921, 612.69651, 631.74498, 669.9~
$ Population             <dbl> 33736494, 327582, 31731688, 3696958, 2~
```
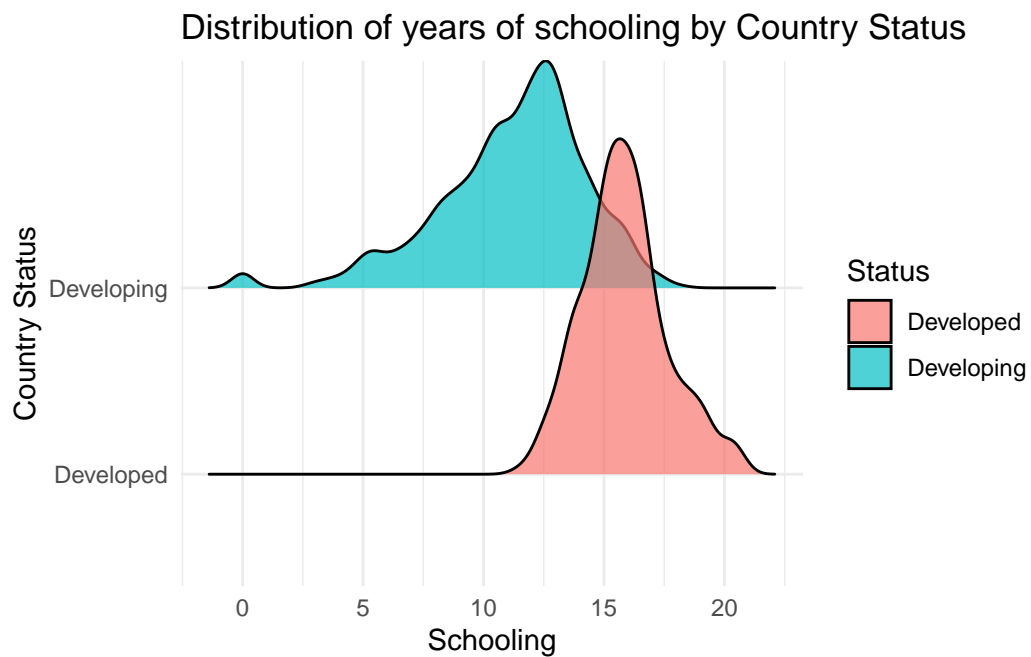
```
$ thinness..1.19.years          <dbl> 17.2, 17.5, 17.7, 17.9, 18.2, 18.4, 18~
$ thinness.5.9.years            <dbl> 17.3, 17.5, 17.7, 18.0, 18.2, 18.4, 18~
$ Income.composition.of.resources <dbl> 0.479, 0.476, 0.470, 0.463, 0.454, 0.4~
$ Schooling                     <dbl> 10.1, 10.0, 9.9, 9.8, 9.5, 9.2, 8.9, 8~
```

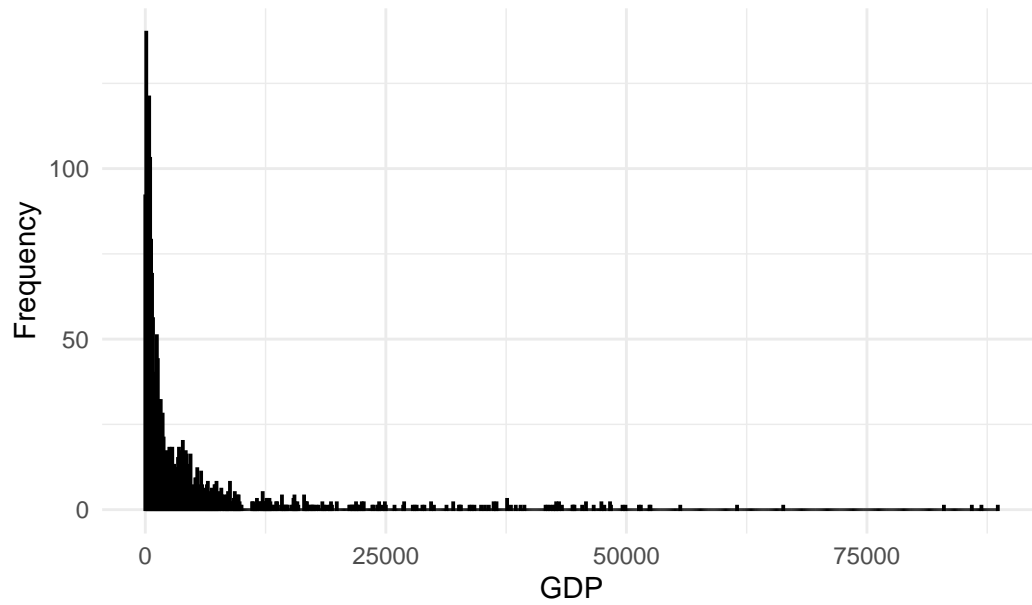**Exploratory Plots:** Dataset 2: Research Question 1

Distribution of Countries by Status


```
Picking joint bandwidth of 0.463
```
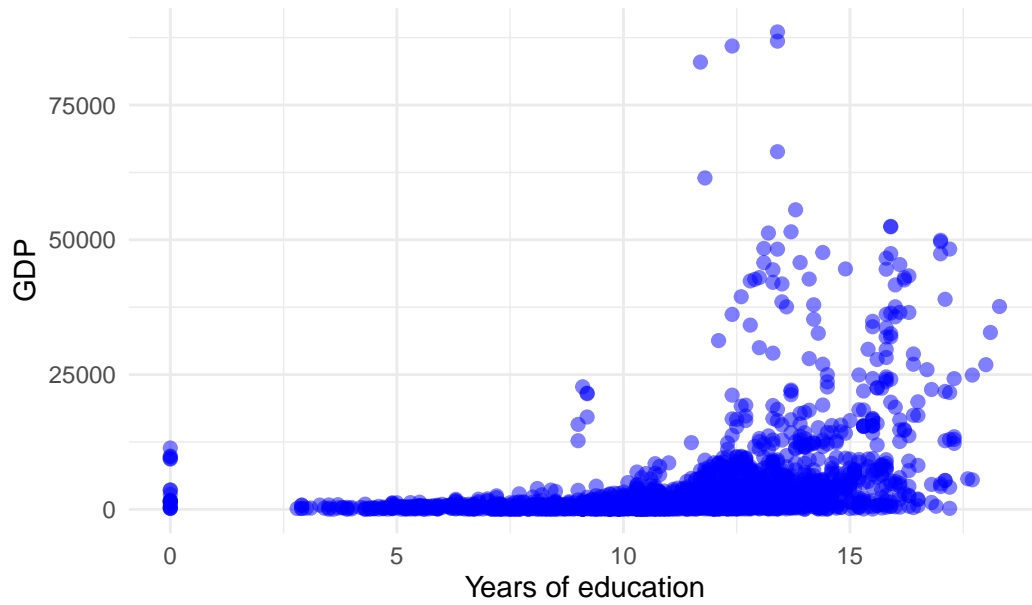
Distribution of years of schooling by Country Status

Distribution of GDP

Relationship Between years of education and GDP

Dataset 3 (optional)

Data source: https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos

Brief description: Each row in the dataset represents data of a female. Specifically, it contains medical, physical, hormonal, and lifestyle information collected for that woman as part of a study on PCOS (Polycystic Ovary Syndrome) and related health metrics.

Research question 1: - What combination of physical characteristics (e.g., BMI, waist-hip ratio), hormonal factors (e.g., AMH, FSH/LH ratio), and symptoms (e.g., pimples, hair loss) significantly predict PCOS status (Y/N)?

- Outcome variable (include the name/description and type of variable): PCOS Status

Categorical (Binary: 1 = Has PCOS, 0 = Does not have PCOS) - Description: Indicates whether the individual has PCOS or not.

Research question 2: How do physical factors and hormonal factors levels influence Hb (hemoglobin) levels?

- Outcome variable (include the name/description and type of variable): Hb (Hemoglobin) Levels

- Type: Continuous (measured in g/dL)

- Description:The concentration of hemoglobin in the blood, which can be an indicator of overall health, anemia, and other medical conditions.

Load the data and provide a `glimpse()`:

```
New names:
* `` -> `...45`


Rows: 541
Columns: 45
$ `Sl. No`            <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14~
$ `Patient File No.`  <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14~
$ `PCOS (Y/N)`        <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
$ `Age (yrs)`         <dbl> 28, 36, 33, 37, 25, 36, 34, 33, 32, 36, 20, 2~
$ `Weight (Kg)`       <dbl> 44.6, 65.0, 68.8, 65.0, 52.0, 74.1, 64.0, 58.~
$ `Height(Cm)`        <dbl> 152.0, 161.5, 165.0, 148.0, 161.0, 165.0, 156~
$ BMI                 <dbl> 19.30000, 24.92116, 25.27089, 29.67495, 20.06~
$ `Blood Group`       <dbl> 15, 15, 11, 13, 11, 15, 11, 13, 11, 15, 15, 1~
$ `Pulse rate(bpm)`   <dbl> 78, 74, 72, 72, 72, 78, 72, 72, 72, 80, 80, 7~
```
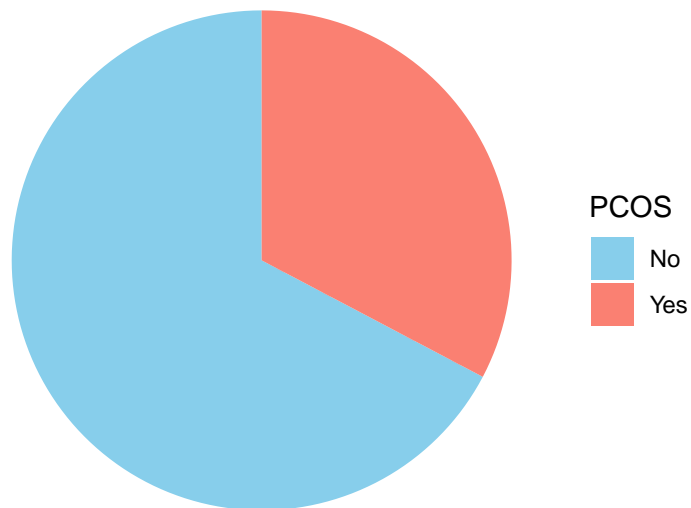
```
$ `RR (breaths/min)`      <dbl> 22, 20, 18, 20, 18, 28, 18, 20, 18, 20, 20, 2~
$ `Hb(g/dl)`              <dbl> 10.48, 11.70, 11.80, 12.00, 10.00, 11.20, 10.~
$ `Cycle(R/I)`           <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 2, 2, 4, 2, 2, ~
$ `Cycle length(days)`   <dbl> 5, 5, 5, 5, 5, 5, 5, 5, 5, 2, 5, 5, 2, 5, 5, ~
$ `Marraige Status (Yrs)` <dbl> 7, 11, 10, 4, 1, 8, 2, 13, 8, 4, 4, 3, 7, 15,~
$ `Pregnant(Y/N)`        <dbl> 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, ~
$ `No. of aborptions`    <dbl> 0, 0, 0, 0, 0, 0, 0, 2, 1, 0, 2, 1, 0, 0, 0, ~
$ `I   beta-HCG(mIU/mL)` <dbl> 1.99, 60.80, 494.08, 1.99, 801.45, 237.97, 1.~
$ `II    beta-HCG(mIU/mL)` <chr> "1.99", "1.99", "494.08", "1.99", "801.45", "~
$ `FSH(mIU/mL)`          <dbl> 7.95, 6.73, 5.54, 8.06, 3.98, 3.24, 2.85, 4.8~
$ `LH(mIU/mL)`           <dbl> 3.68, 1.09, 0.88, 2.36, 0.90, 1.07, 0.31, 3.0~
$ `FSH/LH`               <dbl> 2.160326, 6.174312, 6.295455, 3.415254, 4.422~
$ `Hip(inch)`            <dbl> 36, 38, 40, 42, 37, 44, 39, 44, 39, 40, 39, 3~
$ `Waist(inch)`          <dbl> 30, 32, 36, 36, 30, 38, 33, 38, 35, 38, 35, 3~
$ `Waist:Hip Ratio`      <dbl> 0.8333333, 0.8421053, 0.9000000, 0.8571429, 0~
$ `TSH (mIU/L)`          <dbl> 0.68, 3.16, 2.54, 16.41, 3.57, 1.60, 1.51, 12~
$ `AMH(ng/mL)`           <chr> "2.07", "1.53", "6.63", "1.22", "2.26", "6.74~
$ `PRL(ng/mL)`           <dbl> 45.16, 20.09, 10.52, 36.90, 30.09, 16.18, 26.~
$ `Vit D3 (ng/mL)`       <dbl> 17.10, 61.30, 49.70, 33.40, 43.80, 52.40, 42.~
$ `PRG(ng/mL)`           <dbl> 0.57, 0.97, 0.36, 0.36, 0.38, 0.30, 0.46, 0.2~
$ `RBS(mg/dl)`           <dbl> 92, 92, 84, 76, 84, 76, 93, 91, 116, 125, 108~
$ `Weight gain(Y/N)`     <dbl> 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, ~
$ `hair growth(Y/N)`     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
$ `Skin darkening (Y/N)` <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
$ `Hair loss(Y/N)`       <dbl> 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, ~
$ `Pimples(Y/N)`         <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, ~
$ `Fast food (Y/N)`      <dbl> 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, ~
$ `Reg.Exercise(Y/N)`    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
$ `BP _Systolic (mmHg)`  <dbl> 110, 120, 120, 120, 120, 110, 120, 120, 120, ~
$ `BP _Diastolic (mmHg)` <dbl> 80, 70, 80, 70, 80, 70, 80, 80, 80, 80, 80, 8~
$ `Follicle No. (L)`     <dbl> 3, 3, 13, 2, 3, 9, 6, 7, 5, 1, 7, 4, 15, 3, 4~
$ `Follicle No. (R)`     <dbl> 3, 5, 15, 2, 4, 6, 6, 6, 7, 1, 15, 2, 8, 3, 1~
$ `Avg. F size (L) (mm)` <dbl> 18, 15, 18, 15, 16, 16, 15, 15, 17, 14, 17, 1~
$ `Avg. F size (R) (mm)` <dbl> 18, 14, 20, 14, 14, 20, 16, 18, 17, 17, 20, 1~
$ `Endometrium (mm)`     <dbl> 8.5, 3.7, 10.0, 7.5, 7.0, 8.0, 6.8, 7.1, 4.2,~
$ ...45                  <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```
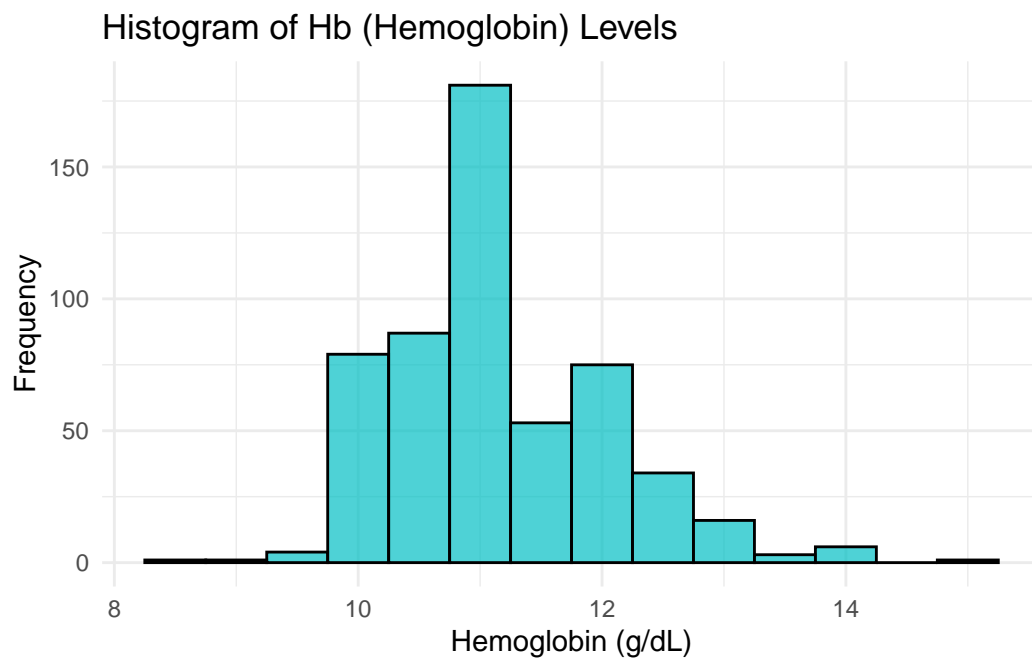
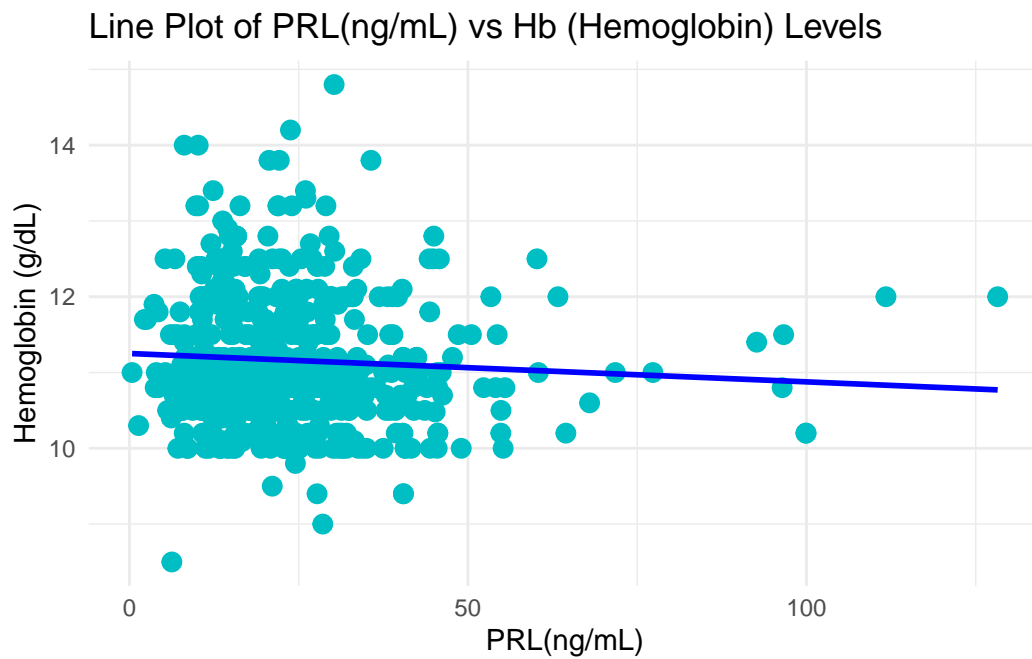**Exploratory Plots:** Dataset 3: Research Question 1

PCOS Distribution

Boxplot of Weight by PCOS Status

Dataset 3: Research Question 2

Histogram of Hb (Hemoglobin) Levels
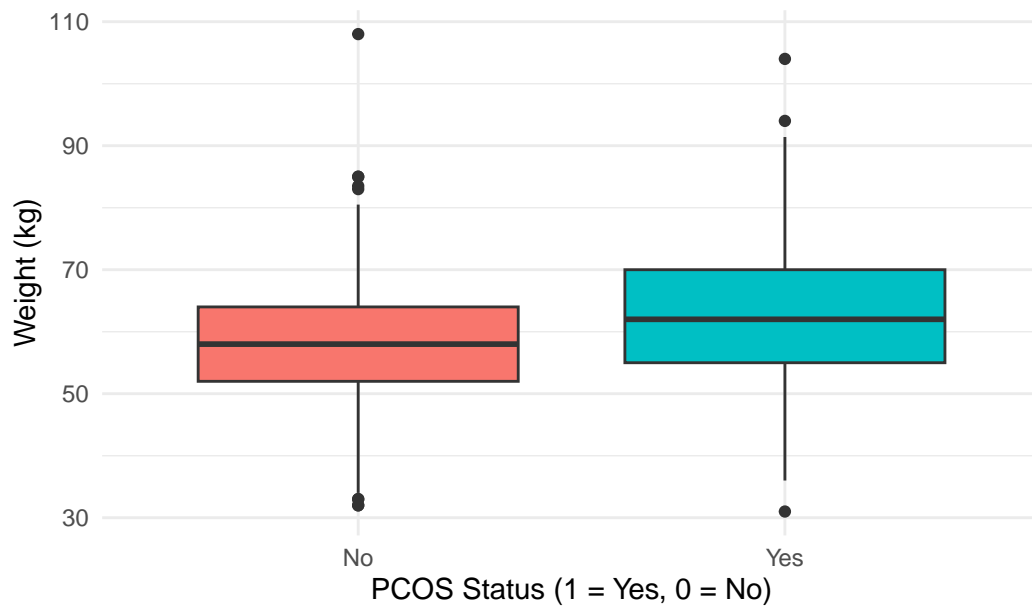
`geom_smooth()` using formula = 'y ~ x'

Line Plot of PRL(ng/mL) vs Hb (Hemoglobin) Levels

Team Charter

When will you meet as a team to work on the project components? Will these meetings be held in person or virtually?

- We will meet virtually (i.e through Zoom) every 7 days in the initial phase. However, near the deadline, we will aim to meet every 2 days and give updates on our individual tasks.

What is your group policy on missing team meetings (e.g., how much advance notice should be provided)?

- Given that we are all juggling multiple projects/deadline, we can miss a meeting provided we give a 24 hour notice and follow-up with what we miss in the meetings and do the work that we are assigned

How will your team communicate (email, Slack, text messages)? What is your policy on appropriate response time (within a certain number of hours? Nights/weekends?)?

- Our main form of communication would be WhatsApp.
- We are not obligated to respond in the night, however if we are required to respond in the day, a half day buffer is acceptable