Milestone2

Mikayla Lamping

3/13/2022

Maternal Risk

Check for potential data quality issues

```
maternal_risk <- read_csv("MaternalHealthRiskDataSet.csv")
```

```
## Rows: 1014 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): RiskLevel
## dbl (6): Age, SystolicBP, DiastolicBP, BS, BodyTemp, HeartRate
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# No missing values
sum(maternal_risk %>%
  is.na())
```

```
## [1] 0
```

Variable definitions from Kaggle: - Age: Age in years when a woman is pregnant. - SystolicBP: Upper value of Blood Pressure in mmHg, another significant attribute during pregnancy. - DiastolicBP: Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy. - BS: Blood glucose levels is in terms of a molar concentration, mmol/L. - HeartRate: A normal resting heart rate in beats per minute. - Risk Level: Predicted Risk Intensity Level during pregnancy considering the previous attribute.

```
summary(maternal_risk)
```

```
##       Age          SystolicBP     DiastolicBP          BS
##  Min.   :10.00   Min.   : 70.0   Min.   : 49.00   Min.   : 6.000
##  1st Qu.:19.00   1st Qu.:100.0   1st Qu.: 65.00   1st Qu.: 6.900
##  Median :26.00   Median :120.0   Median : 80.00   Median : 7.500
##  Mean   :29.87   Mean   :113.2   Mean   : 76.46   Mean   : 8.726
##  3rd Qu.:39.00   3rd Qu.:120.0   3rd Qu.: 90.00   3rd Qu.: 8.000
##  Max.   :70.00   Max.   :160.0   Max.   :100.00   Max.   :19.000
##     BodyTemp        HeartRate     RiskLevel
##  Min.   : 98.00   Min.   : 7.0   Length:1014
##  1st Qu.: 98.00   1st Qu.:70.0   Class :character
##  Median : 98.00   Median :76.0   Mode  :character
##  Mean   : 98.67   Mean   :74.3
##  3rd Qu.: 98.00   3rd Qu.:80.0
##  Max.   :103.00   Max.   :90.0
```

HeartRate seems to have at least one outlier, as the minimum is only 7 but the majority of the points are in the 70-80 bpm range.

Check for duplicate entries

```
maternal_risk %>%
  distinct()
```

```
## # A tibble: 452 x 7
##      Age SystolicBP DiastolicBP    BS BodyTemp HeartRate RiskLevel
##    <dbl>      <dbl>       <dbl> <dbl>    <dbl>     <dbl> <chr>
##  1    25        130          80 15          98        86 high risk
##  2    35        140          90 13          98        70 high risk
##  3    29         90          70  8         100        80 high risk
##  4    30        140          85  7          98        70 high risk
##  5    35        120          60  6.1        98        76 low risk
##  6    23        140          80  7.01       98        70 high risk
##  7    23        130          70  7.01       98        78 mid risk
##  8    35         85          60 11         102        86 high risk
##  9    32        120          90  6.9        98        70 mid risk
## 10    42        130          80 18          98        70 high risk
## # ... with 442 more rows
```

There appear to be many duplicates, as distinct() reduced the number of rows from 1014 to just 452. It does seem that the duplicate entries still represent different people and they just happened to have the same measurements, so these entries should still be included as we proceed with our analysis.

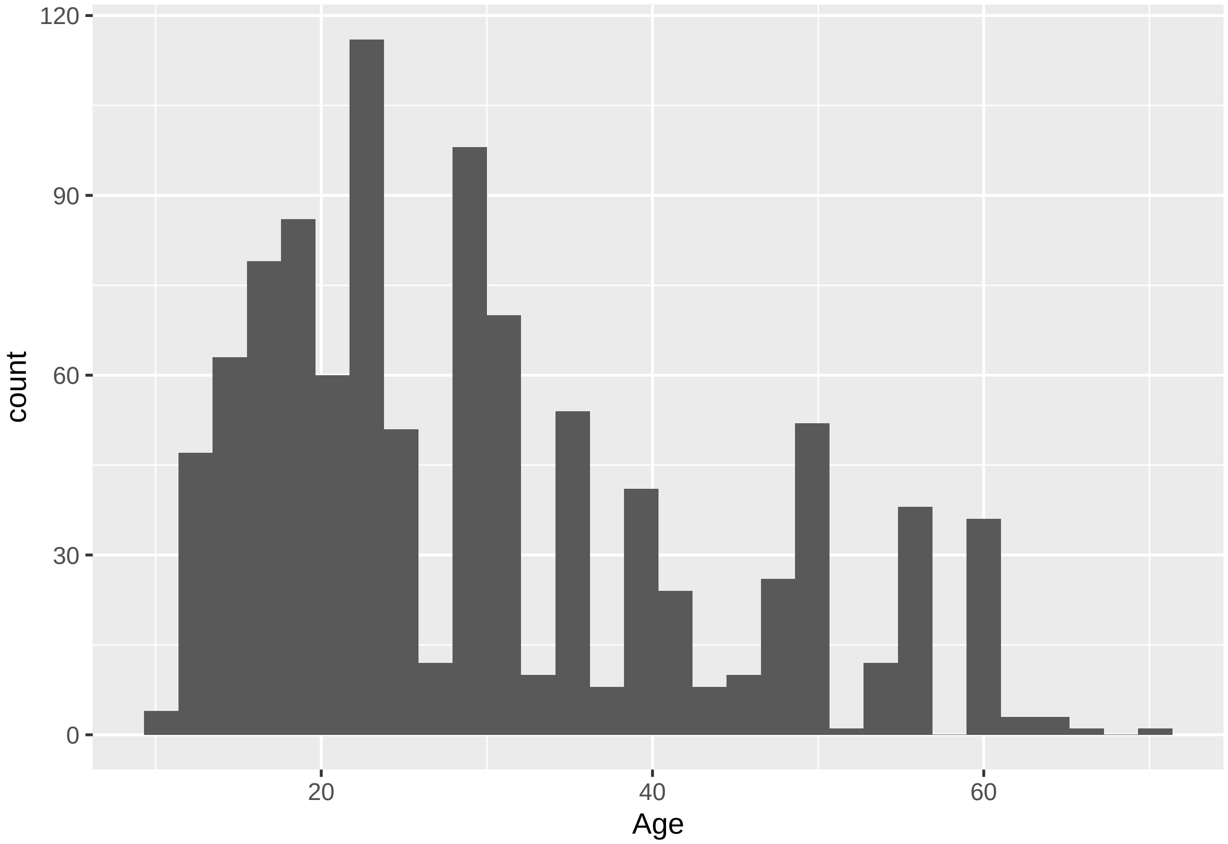
Build intuition about individual variables

```
histogram <- function(colname) {
  maternal_risk %>%
    ggplot() +
    geom_histogram(aes(x = .data[[colname]]))
}
```

Age

```
histogram("Age")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

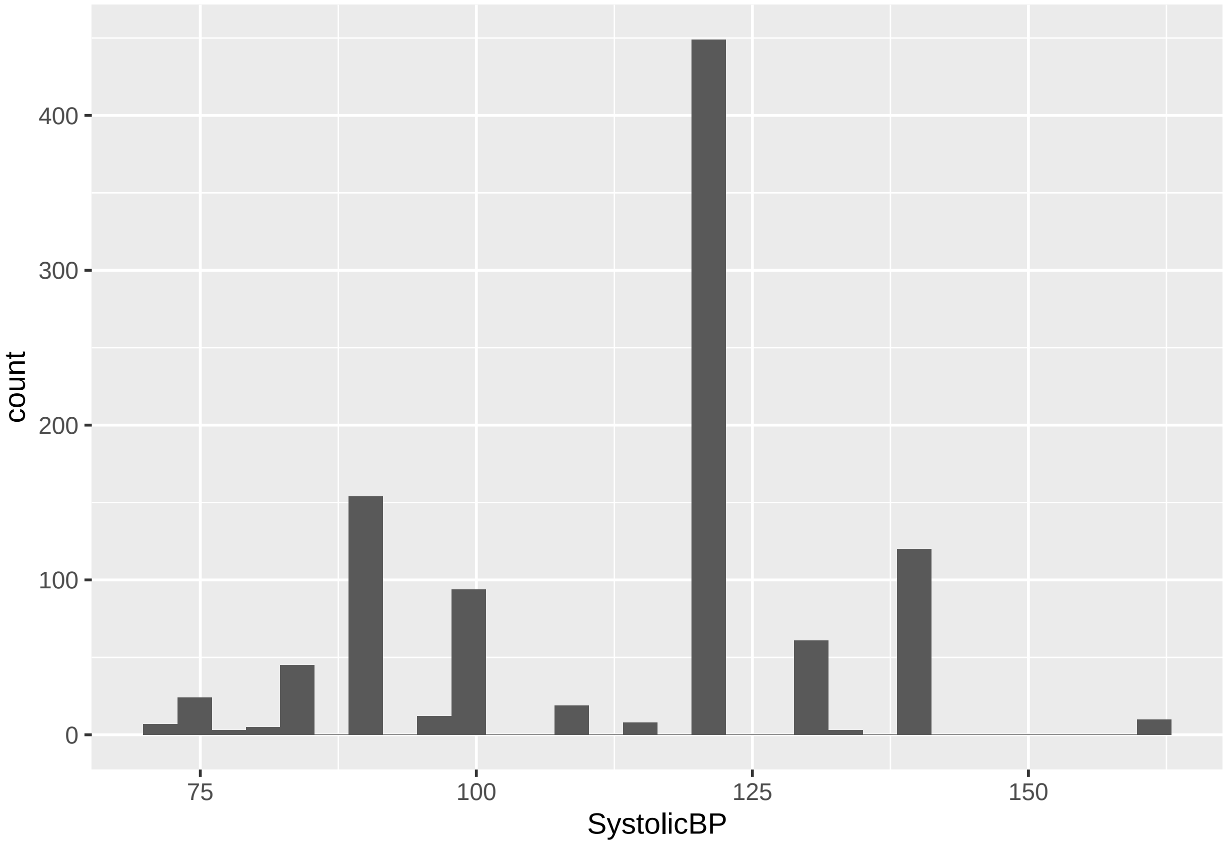


Age ranges from 10 to 70, with the majority between 15 and 25, which is expected.

SystolicBP

```
histogram("SystolicBP")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Small number of entries with SystolicBP above 150. What are their risk levels?

```
maternal_risk %>%
  filter(SystolicBP > 150)
```

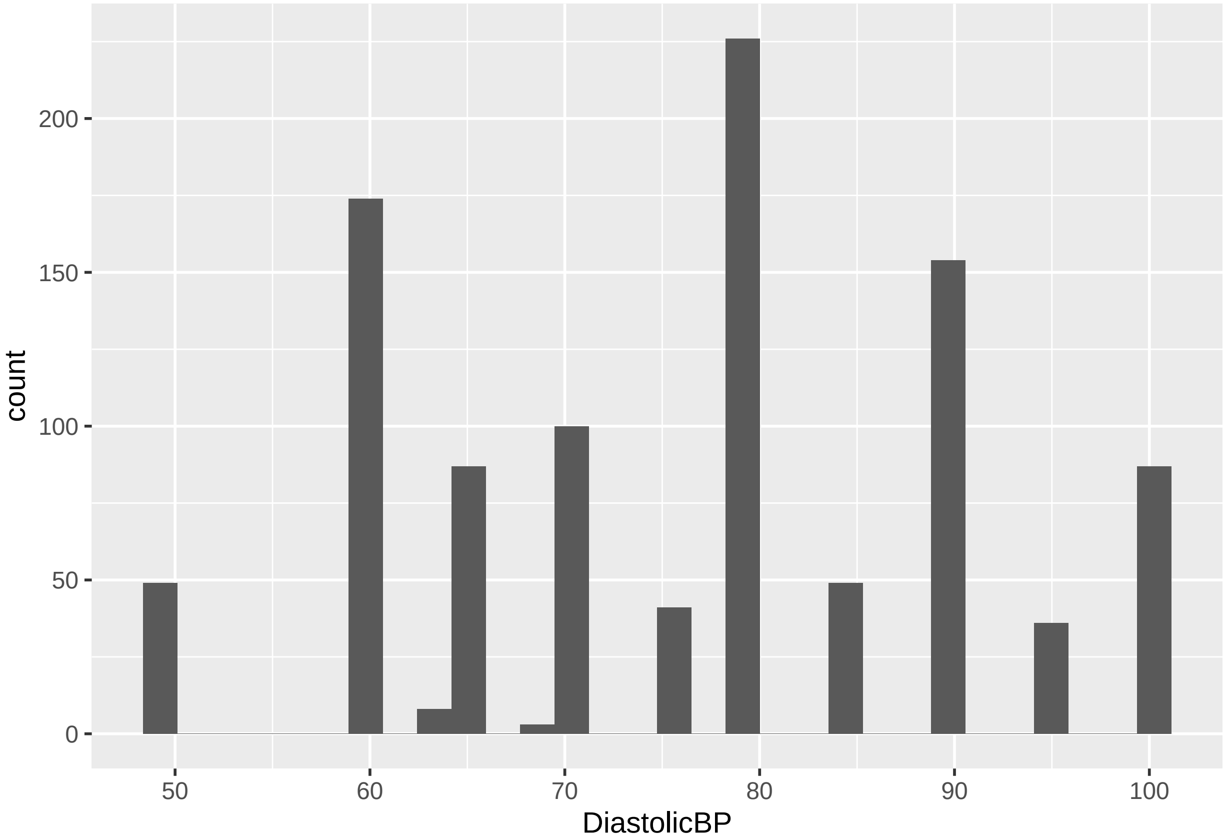
```
## # A tibble: 10 x 7
##      Age SystolicBP DiastolicBP    BS BodyTemp HeartRate RiskLevel
##    <dbl>      <dbl>       <dbl> <dbl>    <dbl>     <dbl> <chr>
##  1    40        160         100    19       98        77 high risk
##  2    40        160         100    19       98        77 high risk
##  3    40        160         100    19       98        77 high risk
##  4    40        160         100    19       98        77 high risk
##  5    40        160         100    19       98        77 high risk
##  6    40        160         100    19       98        77 high risk
##  7    40        160         100    19       98        77 high risk
##  8    40        160         100    19       98        77 high risk
##  9    40        160         100    19       98        77 high risk
## 10    40        160         100    19       98        77 high risk
```

All ten rows are identical and classify each as high risk, which is expected given that these entries are outliers.

DiastolicBP

```
histogram("DiastolicBP")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Fairly evenly distributed. Check under 50 group.

```
maternal_risk %>%
  filter(DiastolicBP < 50)
```

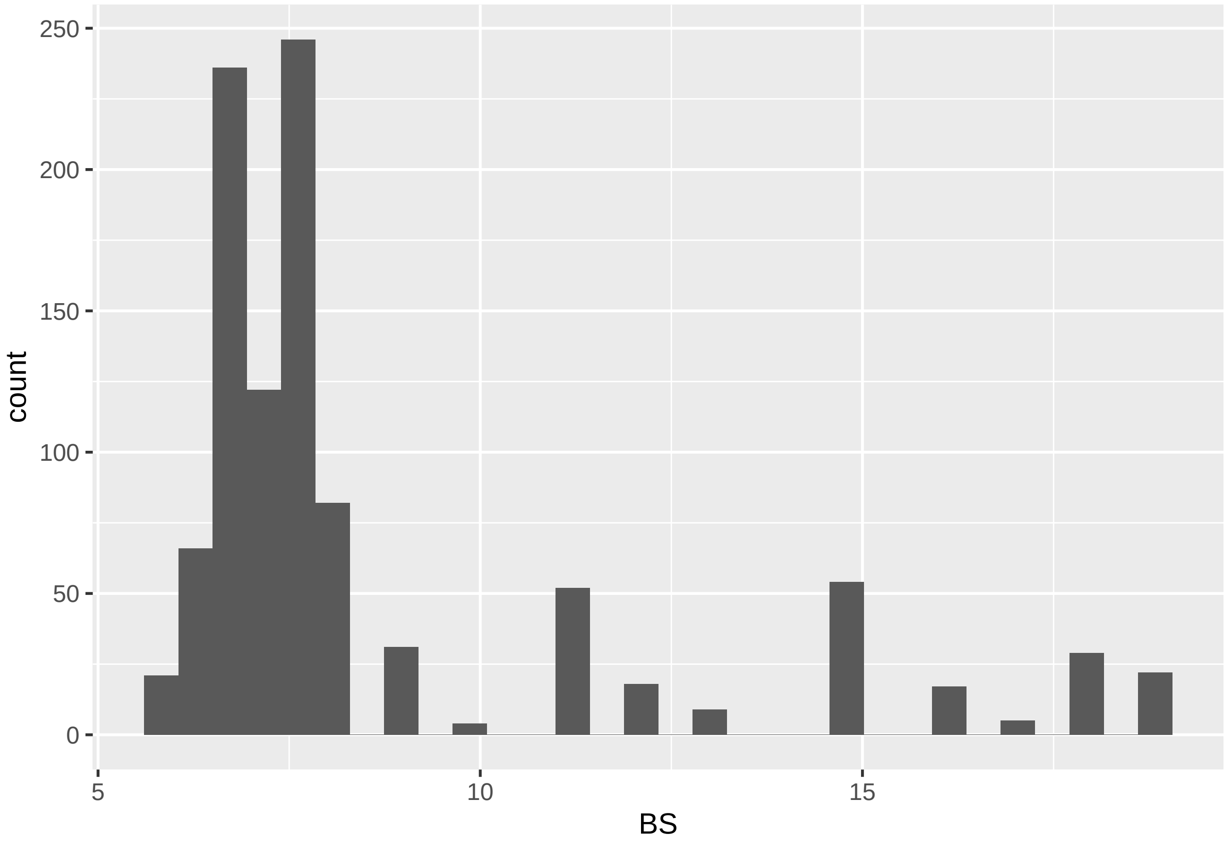
```
## # A tibble: 25 x 7
##      Age SystolicBP DiastolicBP    BS BodyTemp HeartRate RiskLevel
##    <dbl>      <dbl>       <dbl> <dbl>    <dbl>     <dbl> <chr>    
##  1    15         76          49   7.5       98        77 low risk 
##  2    15         76          49   6.4       98        77 low risk 
##  3    15         76          49   7.5       98        77 low risk 
##  4    15         76          49   6.4       98        77 low risk 
##  5    15         75          49   7.7       98        77 low risk 
##  6    15         76          49   7.8       98        77 low risk 
##  7    15         76          49   6.8       98        77 low risk 
##  8    15         76          49   6.8       98        77 low risk 
##  9    15         76          49   7.9       98        77 low risk 
## 10    15         76          49   7.9       98        77 low risk 
## # ... with 15 more rows
```

25 entries with Diastolic blood pressure less than 50 (all 49). All are classified as low risk, so upon initial exploration, these low DiastolicBP outliers don't seem to increase risk level.

Blood Sugar

```
histogram("BS")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The majority are under 10, but there are many entries spread between 10 and 20. Do these have higher risk level?

```
maternal_risk %>%
  filter(BS > 10) %>%
  group_by(RiskLevel) %>%
  summarise(count = n())
```

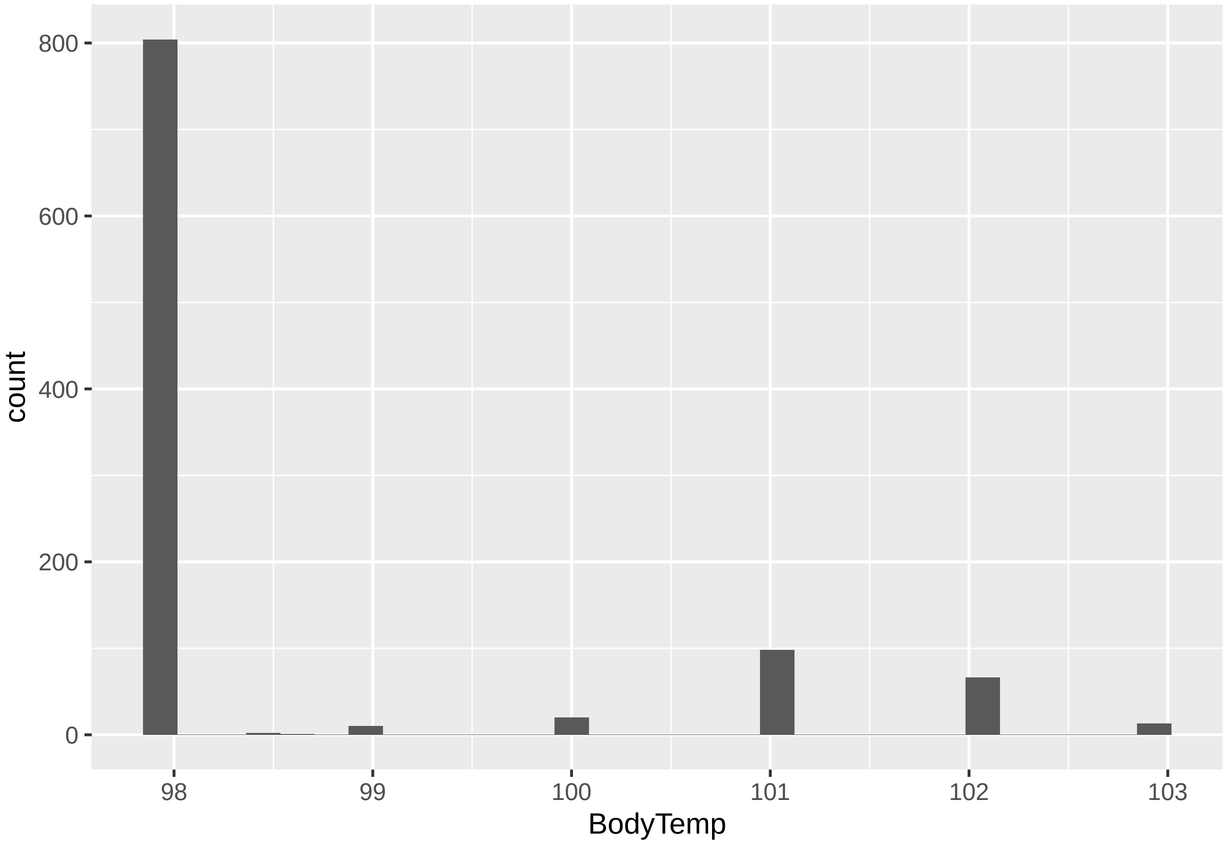
```
## # A tibble: 3 x 2
##   RiskLevel count
##   <chr>     <int>
## 1 high risk   170
## 2 low risk      4
## 3 mid risk     32
```

Yes, the majority of the entries with a blood sugar level greater than 10 are classified as high risk. This is a factor we should explore further, as this seems to indicate that high values of BP correlate with higher risk level.

Body Temperature

```
histogram("BodyTemp")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The majority are in the normal temperature range (~98-99).

```
maternal_risk %>%
  filter(BodyTemp > 100) %>%
  group_by(RiskLevel) %>%
  summarise(count = n())
```

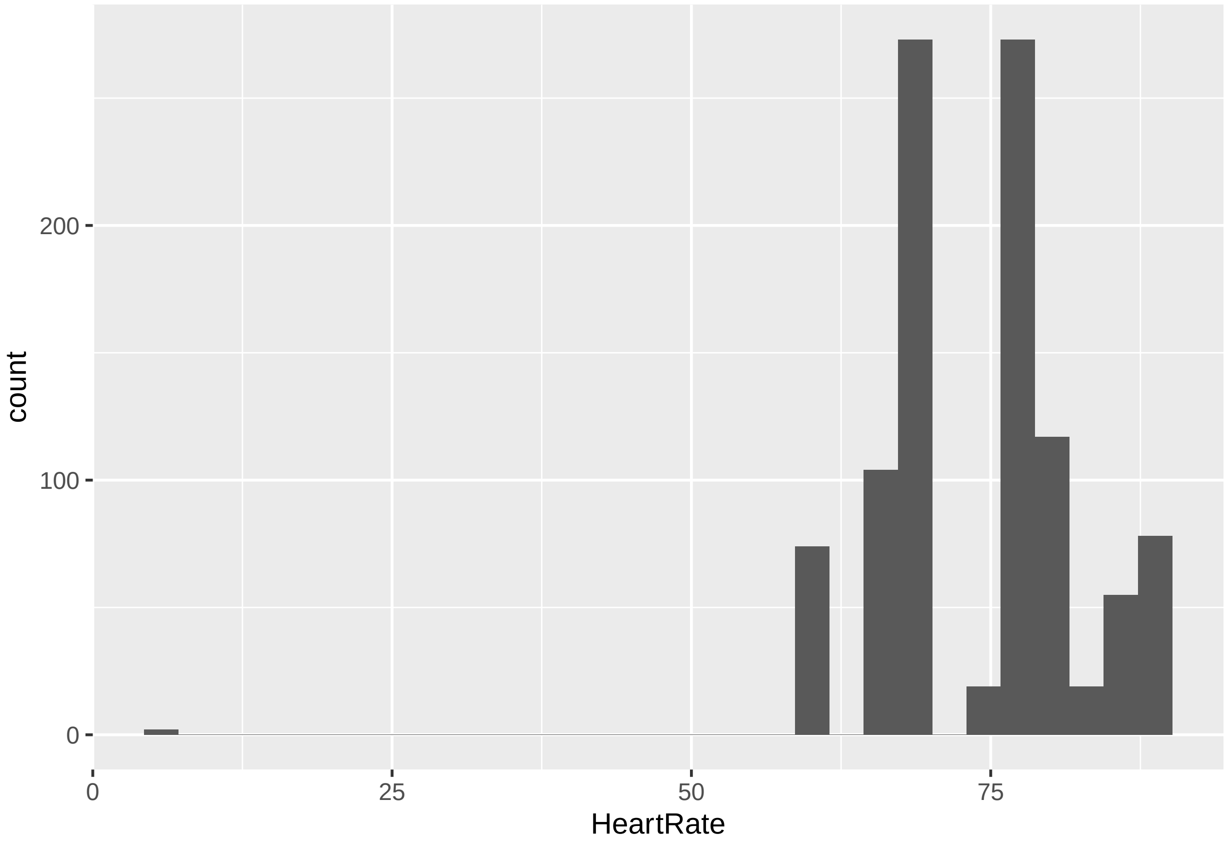
```
## # A tibble: 3 x 2
##   RiskLevel count
##   <chr>     <int>
## 1 high risk    65
## 2 low risk     37
## 3 mid risk     75
```

There doesn't seem to be a super strong correlation, but there are more mid to high risk entries than low risk in the group of higher body temperatures.

HeartRate

```
histogram("HeartRate")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



There is a major outlier, with heart rate less than 10.

```
maternal_risk %>%
  filter(HeartRate < 25)
```

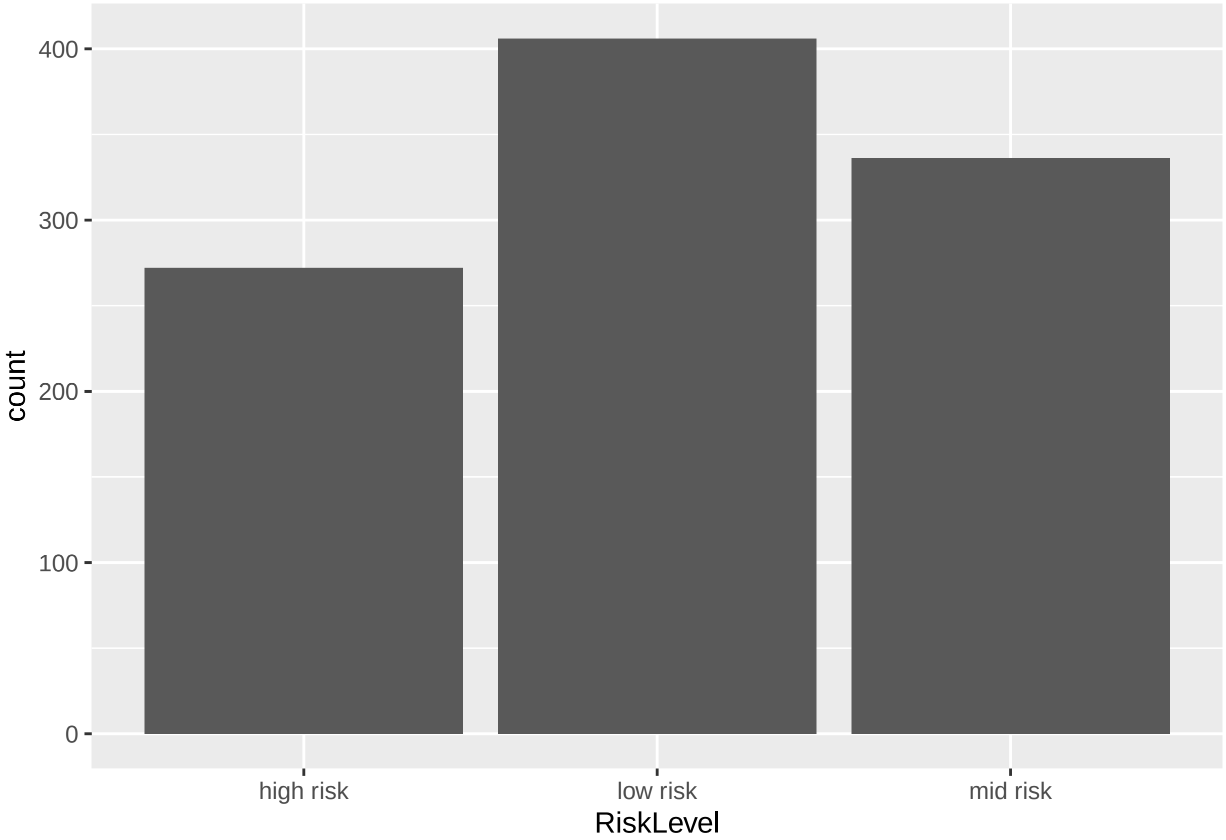
```
## # A tibble: 2 x 7
##     Age SystolicBP DiastolicBP    BS BodyTemp HeartRate RiskLevel
##   <dbl>      <dbl>       <dbl> <dbl>    <dbl>     <dbl> <chr>    
## 1    16        120          75   7.9       98         7 low risk 
## 2    16        120          75   7.9       98         7 low risk
```

HeartRate may have been entered incorrectly? This person is classified as low risk but has a heart rate of just 7, while the majority of heart rates in this dataset are around 70 bpm.

Risk Level

```
# Plot discrete histogram (RiskLevel)
maternal_risk %>%
  ggplot() +
  geom_histogram(aes(x = RiskLevel), stat = "count")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Risk level mainly low but fairly evenly dispersed.

```
maternal_risk %>%
  count(RiskLevel)
```

```
## # A tibble: 3 x 2
##   RiskLevel     n
##   <chr>     <int>
## 1 high risk   272
## 2 low risk    406
## 3 mid risk    336
```

The order of the levels isn't intuitive - modify this in next step.

Prepare processed data

RiskLevel in order that doesn't make sense for plotting. Update factor to have levels increase from low to high.

```
risks <- (c("low risk", "mid risk", "high risk"))

maternal_risk2 <- maternal_risk %>%
  mutate(RiskLevel = factor(RiskLevel, levels = risks))
```

Global Mortality

Check for potential data quality issues

```
global <- read_csv("global_mortality.csv")
```

```
## Rows: 6156 Columns: 35
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): country, country_code
## dbl (33): year, Cardiovascular diseases (%), Cancers (%), Respiratory diseas...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# no missing values in data we care about
sum(global %>%
      select(c(`Neonatal deaths (%)`, `Maternal deaths (%)`)) %>%
      is.na())
```

```
## [1] 0
```

Key Variables: - Maternal deaths (%): Percentage of all deaths in a given country & year where the cause was linked to maternal/childbirth complications - Neonatal deaths (%): Percentage of all deaths in a given country & year where death occurred within the first 28 days of life

```
summary(global[c(1, 10, 25)])
```

```
##    country           Neonatal deaths (%) Maternal deaths (%)
##  Length:6156        Min.   : 0.04071    Min.   :0.00188    
##  Class :character   1st Qu.: 0.68559    1st Qu.:0.03230    
##  Mode  :character   Median : 3.89183    Median :0.23696    
##                     Mean   : 4.56666    Mean   :0.58591    
##                     3rd Qu.: 7.74003    3rd Qu.:1.00166    
##                     Max.   :17.80683    Max.   :3.41435    
```

Check for duplicate entries

```
# same size as original dataset, no duplicates
nrow(global %>% distinct())
```

```
## [1] 6156
```

```
length(global %>% distinct())
```

```
## [1] 35
```

Prepare processed data

```
# keep most recent year
global <- global %>%
  filter(year == 2016)
```

Build intuition about individual variables

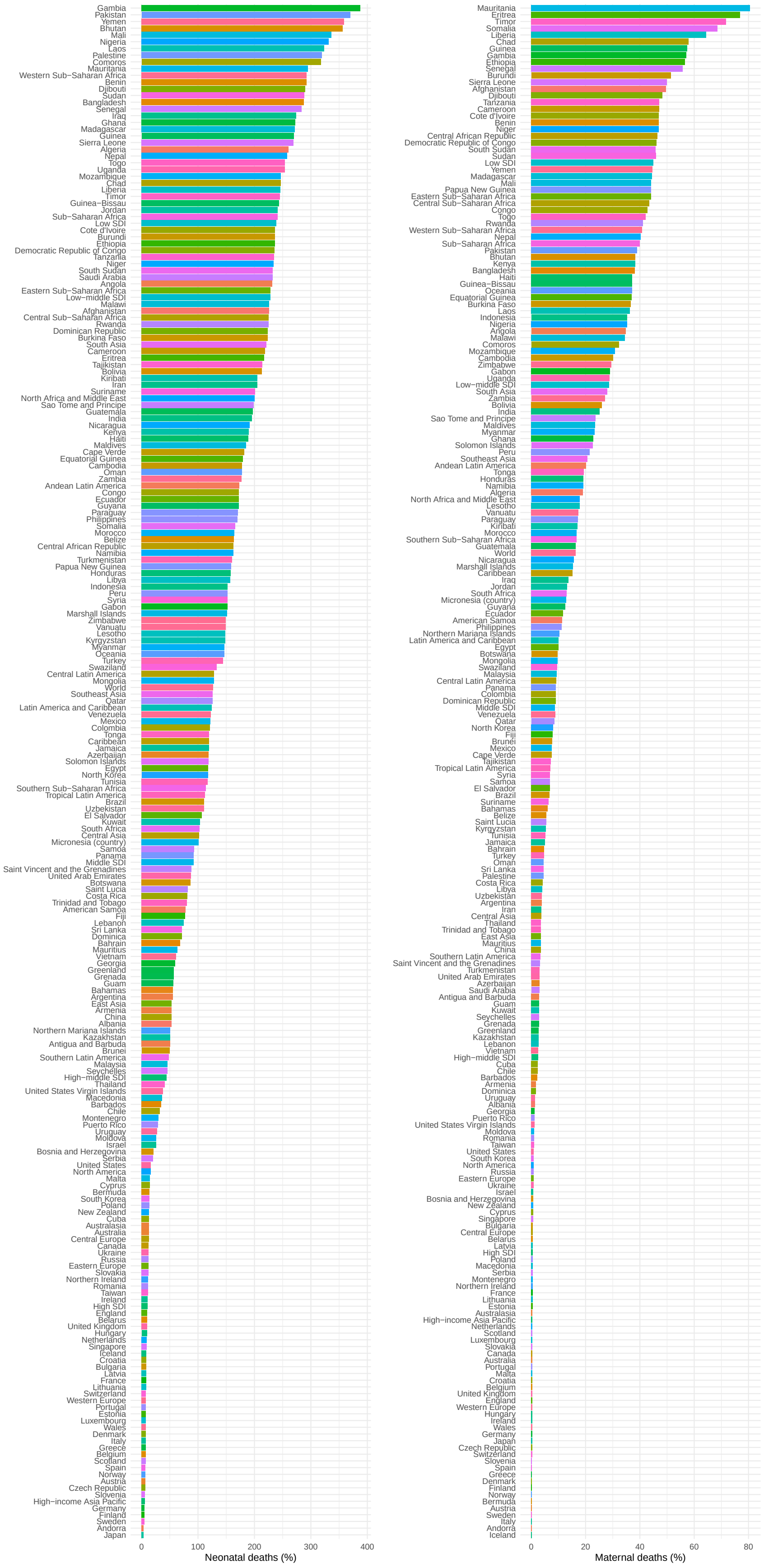
```
global <- read_csv("global_mortality.csv")
```

```
## Rows: 6156 Columns: 35
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): country, country_code
## dbl (33): year, Cardiovascular diseases (%), Cancers (%), Respiratory diseas...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
countries1 <- ggplot(global) +
  geom_col(aes(`Neonatal deaths (%)`, reorder(country, `Neonatal deaths (%)`), fill = country)) +
  ylab(label = "") +
  theme(axis.text = element_text(size = 2)) +
  theme_minimal() +
  guides(fill="none")

countries2 <- ggplot(global) +
  geom_col(aes(`Maternal deaths (%)`, reorder(country, `Maternal deaths (%)`), fill = country)) +
  ylab(label = "") +
  theme(axis.text = element_text(size = 2)) +
  theme_minimal() +
  guides(fill="none")

countries1 + countries2
```


Gambia
Pakistan
Yemen
Bhutan
Mali
Nigeria
Laos
Palestine
Comoros
Mauritania
Western Sub−Saharan Africa
Benin
Djibouti
Sudan
Bangladesh
Senegal
Iraq
Ghana
Madagascar
Guinea
Sierra Leone
Algeria
Nepal
Togo
Uganda
Mozambique
Chad
Liberia
Timor
Guinea−Bissau
Jordan
Sub−Saharan Africa
Low SDI
Cote d'Ivoire
Burundi
Ethiopia
Democratic Republic of Congo
Tanzania
Niger
South Sudan
Saudi Arabia
Angola
Eastern Sub−Saharan Africa
Low−middle SDI
Malawi
Afghanistan
Central Sub−Saharan Africa
Rwanda
Dominican Republic
Burkina Faso
South Asia
Cameroon
Eritrea
Tajikistan
Bolivia
Kiribati
Iran
Suriname
North Africa and Middle East
Sao Tome and Principe
Guatemala
India
Nicaragua
Kenya
Haiti
Maldives
Cape Verde
Equatorial Guinea
Cambodia
Oman
Zambia
Andean Latin America
Congo
Ecuador
Guyana
Paraguay
Philippines
Somalia
Morocco
Belize
Central African Republic
Namibia
Turkmenistan
Papua New Guinea
Honduras
Libya
Indonesia
Peru
Syria
Gabon
Marshall Islands
Zimbabwe
Vanuatu
Lesotho
Kyrgyzstan
Myanmar
Oceania
Turkey
Swaziland
Central Latin America
Mongolia
World
Southeast Asia
Qatar
Latin America and Caribbean
Venezuela
Mexico
Colombia
Tonga
Caribbean
Jamaica
Azerbaijan
Solomon Islands
Egypt
North Korea
Tunisia
Southern Sub−Saharan Africa
Tropical Latin America
Brazil
Uzbekistan
El Salvador
Kuwait
South Africa
Central Asia
Micronesia (country)
Samoa
Panama
Middle SDI
Saint Vincent and the Grenadines
United Arab Emirates
Botswana
Saint Lucia
Costa Rica
Trinidad and Tobago
American Samoa
Fiji
Lebanon
Sri Lanka
Dominica
Bahrain
Mauritius
Vietnam
Georgia
Greenland
Grenada
Guam
Bahamas
Argentina
East Asia
Armenia
China
Albania
Northern Mariana Islands
Kazakhstan
Antigua and Barbuda
Brunei
Southern Latin America
Malaysia
Seychelles
High−middle SDI
Thailand
United States Virgin Islands
Macedonia
Barbados
Chile
Montenegro
Puerto Rico
Uruguay
Moldova
Israel
Bosnia and Herzegovina
Serbia
United States
North America
Malta
Cyprus
Bermuda
South Korea
Poland
New Zealand
Cuba
Australasia
Australia
Central Europe
Canada
Ukraine
Russia
Eastern Europe
Slovakia
Northern Ireland
Romania
Taiwan
Ireland
High SDI
England
Belarus
United Kingdom
Hungary
Netherlands
Singapore
Iceland
Croatia
Bulgaria
Latvia
France
Lithuania
Switzerland
Western Europe
Portugal
Estonia
Luxembourg
Wales
Denmark
Italy
Greece
Belgium
Scotland
Spain
Norway
Austria
Czech Republic
Slovenia
High−income Asia Pacific
Germany
Finland
Sweden
Andorra
Japan
0
100
200
300
400
Neonatal deaths (%)
Mauritania
Eritrea
Timor
Somalia
Liberia
Chad
Guinea
Gambia
Ethiopia
Senegal
Burundi
Sierra Leone
Afghanistan
Djibouti
Tanzania
Cameroon
Cote d'Ivoire
Benin
Niger
Central African Republic
Democratic Republic of Congo
South Sudan
Sudan
Low SDI
Yemen
Madagascar
Mali
Papua New Guinea
Eastern Sub−Saharan Africa
Central Sub−Saharan Africa
Congo
Togo
Rwanda
Western Sub−Saharan Africa
Nepal
Sub−Saharan Africa
Pakistan
Bhutan
Kenya
Bangladesh
Haiti
Guinea−Bissau
Oceania
Equatorial Guinea
Burkina Faso
Laos
Indonesia
Nigeria
Angola
Malawi
Comoros
Mozambique
Cambodia
Zimbabwe
Gabon
Uganda
Low−middle SDI
South Asia
Zambia
Bolivia
India
Sao Tome and Principe
Maldives
Myanmar
Ghana
Solomon Islands
Peru
Southeast Asia
Andean Latin America
Tonga
Honduras
Namibia
Algeria
North Africa and Middle East
Lesotho
Vanuatu
Paraguay
Kiribati
Morocco
Southern Sub−Saharan Africa
Guatemala
World
Nicaragua
Marshall Islands
Caribbean
Iraq
Jordan
South Africa
Micronesia (country)
Guyana
Ecuador
American Samoa
Philippines
Northern Mariana Islands
Latin America and Caribbean
Egypt
Botswana
Mongolia
Swaziland
Malaysia
Central Latin America
Panama
Colombia
Dominican Republic
Middle SDI
Venezuela
Qatar
North Korea
Fiji
Brunei
Mexico
Cape Verde
Tajikistan
Tropical Latin America
Syria
Samoa
El Salvador
Brazil
Suriname
Bahamas
Belize
Saint Lucia
Kyrgyzstan
Tunisia
Jamaica
Bahrain
Turkey
Oman
Sri Lanka
Palestine
Costa Rica
Libya
Uzbekistan
Argentina
Iran
Central Asia
Thailand
Trinidad and Tobago
East Asia
Mauritius
China
Southern Latin America
Saint Vincent and the Grenadines
Turkmenistan
United Arab Emirates
Azerbaijan
Saudi Arabia
Antigua and Barbuda
Guam
Kuwait
Seychelles
Grenada
Greenland
Kazakhstan
Lebanon
Vietnam
High−middle SDI
Cuba
Chile
Barbados
Armenia
Dominica
Uruguay
Albania
Georgia
Puerto Rico
United States Virgin Islands
Moldova
Romania
Taiwan
United States
South Korea
North America
Russia
Eastern Europe
Ukraine
Israel
Bosnia and Herzegovina
New Zealand
Cyprus
Singapore
Bulgaria
Central Europe
Belarus
Latvia
High SDI
Poland
Macedonia
Serbia
Montenegro
Northern Ireland
France
Lithuania
Estonia
Australasia
High−income Asia Pacific
Netherlands
Scotland
Luxembourg
Slovakia
Canada
Australia
Portugal
Malta
Croatia
Belgium
United Kingdom
England
Western Europe
Hungary
Ireland
Wales
Germany
Japan
Czech Republic
Switzerland
Slovenia
Spain
Greece
Denmark
Finland
Norway
Bermuda
Austria
Sweden
Italy
Andorra
Iceland
0
20
40
60
80
Maternal deaths (%)

Interventions & Maternal Outcomes

Check for potential data quality issues

The following .csv files were hand-generated from the cited paper (more details in the write-up).

```
interventions <- read_csv("interventions.csv")
```

```
## Rows: 3 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): setting, caregiver
## dbl (12): external_tocometer, fetal_scalp_electrode, amniotomy, oxytocin, ni...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
outcomes <- read_csv("maternal_outcomes.csv")
```

```
## Rows: 3 Columns: 20
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): setting, caregiver
## dbl (18): prolapsed_cord, uterine_rupture, postpartum_hemorrhage, blood_tran...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# no missing values
sum(interventions %>%
      is.na())
```

```
## [1] 0
```

```
sum(outcomes %>%
      is.na())
```

```
## [1] 0
```

“interventions” Variables:

Number of pregnant people who had…

- external_tocomoter: external electronic fetal monitoring
- fetal_scalp_electrode: internal electronic fetal monitoring
- amniotomy: purposeful puncturing of the amniotic sac
- oxytocin: drug used to induce labor, brand name Pitocin
- nitrous_oxide: “laughing gas”
- epidural: numbing prodeduce for the lower part of the body
- narcotic: pain-relieving drugs
- spontaneous_vaginal: vaginal birth without forceps or other mechanical interventions
- assisted_vaginal: vaginal birth with mechanical assistance
- cesarean: birth via cesarean-section surgery
- episiotomy: cutting of the perineum for vaginal birth

total: total number in each birth setting/caregiver (same as below)

"outcomes" variables:

Number of pregnant people who experienced…

- prolapsed_cord: a prolapsed umbilical cord
- uterine_rupture: a ruputure of the uterus
- postpartum_hemorrhage: hemorrhaging after giving birth
- blood_transfusion: requiring a blood transfusion
- obstetric_shock: shock, usually caused by excessive blood loss
- death
- placenta_removal: manual removal of the placenta (rather than natural expulsion)
- uterine_prolapse: a prolapsed uterus
- pyrexia: fever
- uti: urinary tract infection
- puerpural_fever: prolonged fever immediately after giving birth
- wound_infection: infection at the site of a wound
- no_tear: no tearing from birth
- first_second_degree: first or second degree tears
- third_fourth_degree: third or fourth degree tears
- unknown_degree: tearing of unknown degree
- cervical_tear: tearing of the cervix

total: total number in each birth setting/caregiver (same as above)

```
summary(interventions)
```

```
##    setting           caregiver         external_tocometer fetal_scalp_electrode
##  Length:3           Length:3           Min.   : 389       Min.   : 60.0
##  Class :character   Class :character   1st Qu.:1180       1st Qu.:153.5
##  Mode  :character   Mode  :character   Median :1970       Median :247.0
##                                        Mean   :2174       Mean   :285.0
##                                        3rd Qu.:3067       3rd Qu.:397.5
##                                        Max.   :4164       Max.   :548.0
##    amniotomy       oxytocin     nitrous_oxide     epidural
##  Min.   : 560   Min.   :172.0   Min.   : 199   Min.   : 224.0
##  1st Qu.:1039   1st Qu.:387.5   1st Qu.: 882   1st Qu.: 562.5
##  Median :1518   Median :603.0   Median :1565   Median : 901.0
##  Mean   :1397   Mean   :585.3   Mean   :1550   Mean   : 870.7
##  3rd Qu.:1815   3rd Qu.:792.0   3rd Qu.:2226   3rd Qu.:1194.0
##  Max.   :2112   Max.   :981.0   Max.   :2887   Max.   :1487.0
##     narcotic      spontaneous_vaginal assisted_vaginal    cesarean
##  Min.   : 122.0   Min.   :2605        Min.   : 86.0    Min.   :208.0
##  1st Qu.: 417.5   1st Qu.:3258        1st Qu.:215.0    1st Qu.:353.0
##  Median : 713.0   Median :3910        Median :344.0    Median :498.0
##  Mean   : 904.0   Mean   :3507        Mean   :388.7    Mean   :431.3
##  3rd Qu.:1295.0   3rd Qu.:3958        3rd Qu.:540.0    3rd Qu.:543.0
##  Max.   :1877.0   Max.   :4007        Max.   :736.0    Max.   :588.0
##    episiotomy        total     
##  Min.   : 84.0   Min.   :2899  
##  1st Qu.:186.5   1st Qu.:3826  
##  Median :289.0   Median :4752  
##  Mean   :391.0   Mean   :4327  
##  3rd Qu.:544.5   3rd Qu.:5042  
##  Max.   :800.0   Max.   :5331  
```

```
summary(outcomes)
```

```
##    setting           caregiver         prolapsed_cord  uterine_rupture
##  Length:3           Length:3           Min.   :2.000   Min.   :0.0000  
##  Class :character   Class :character   1st Qu.:4.000   1st Qu.:0.0000  
##  Mode  :character   Mode  :character   Median :6.000   Median :0.0000  
##                                        Mean   :5.667   Mean   :0.6667  
##                                        3rd Qu.:7.500   3rd Qu.:1.0000  
##                                        Max.   :9.000   Max.   :2.0000  
##  postpartum_hemorrhage blood_transfusion obstetric_shock     death  
##  Min.   :110.0         Min.   : 2.0      Min.   :1       Min.   :0  
##  1st Qu.:197.5         1st Qu.: 6.0      1st Qu.:1       1st Qu.:0  
##  Median :285.0         Median :10.0      Median :1       Median :0  
##  Mean   :250.7         Mean   : 9.0      Mean   :1       Mean   :0  
##  3rd Qu.:321.0         3rd Qu.:12.5      3rd Qu.:1       3rd Qu.:0  
##  Max.   :357.0         Max.   :15.0      Max.   :1       Max.   :0  
##  placenta_removal uterine_prolapse    pyrexia           uti     
##  Min.   :28.00    Min.   :1.000    Min.   : 19.00   Min.   :0.0  
##  1st Qu.:56.50    1st Qu.:1.000    1st Qu.: 43.50   1st Qu.:0.5  
##  Median :85.00    Median :1.000    Median : 68.00   Median :1.0  
##  Mean   :67.67    Mean   :1.333    Mean   : 80.33   Mean   :2.0  
##  3rd Qu.:87.50    3rd Qu.:1.500    3rd Qu.:111.00   3rd Qu.:3.0  
##  Max.   :90.00    Max.   :2.000    Max.   :154.00   Max.   :5.0  
##  puerpural_fever wound_infection    no_tear     first_second_degree
##  Min.   :1.0     Min.   : 0.0    Min.   :1578   Min.   :1262       
##  1st Qu.:2.5     1st Qu.: 5.5    1st Qu.:1884   1st Qu.:1824       
##  Median :4.0     Median :11.0    Median :2189   Median :2387       
##  Mean   :4.0     Mean   : 9.0    Mean   :2019   Mean   :2162       
##  3rd Qu.:5.5     3rd Qu.:13.5    3rd Qu.:2240   3rd Qu.:2612       
##  Max.   :7.0     Max.   :16.0    Max.   :2291   Max.   :2836       
##  third_fourth_degree unknown_degree  cervical_tear       total     
##  Min.   : 34.0       Min.   :21.00   Min.   :2.000   Min.   :2899  
##  1st Qu.: 85.5       1st Qu.:23.00   1st Qu.:3.000   1st Qu.:3826  
##  Median :137.0       Median :25.00   Median :4.000   Median :4752  
##  Mean   :118.0       Mean   :28.33   Mean   :3.667   Mean   :4327  
##  3rd Qu.:160.0       3rd Qu.:32.00   3rd Qu.:4.500   3rd Qu.:5042  
##  Max.   :183.0       Max.   :39.00   Max.   :5.000   Max.   :5331  
```

Some variables have a very low sample size/count, which is not as useful for visualization.

Check for duplicate entries

As these two .csv's were made by hand (and are very small overall), we verify there are no duplicate entries.

Prepare processed data

```
interventions_ratios <- interventions %>%
  mutate_at(vars(-c(setting, caregiver, total)), funs(./total)) %>%
  unite("birth_plan", c(setting, caregiver), sep = ", ")
```

```
## Warning: `funs()` was deprecated in dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
outcomes_ratios <- outcomes %>%
  mutate_at(vars(-c(setting, caregiver, total)), funs(./total)) %>%
  unite("birth_plan", c(setting, caregiver), sep = ", ")
```

Build intuition about individual variables

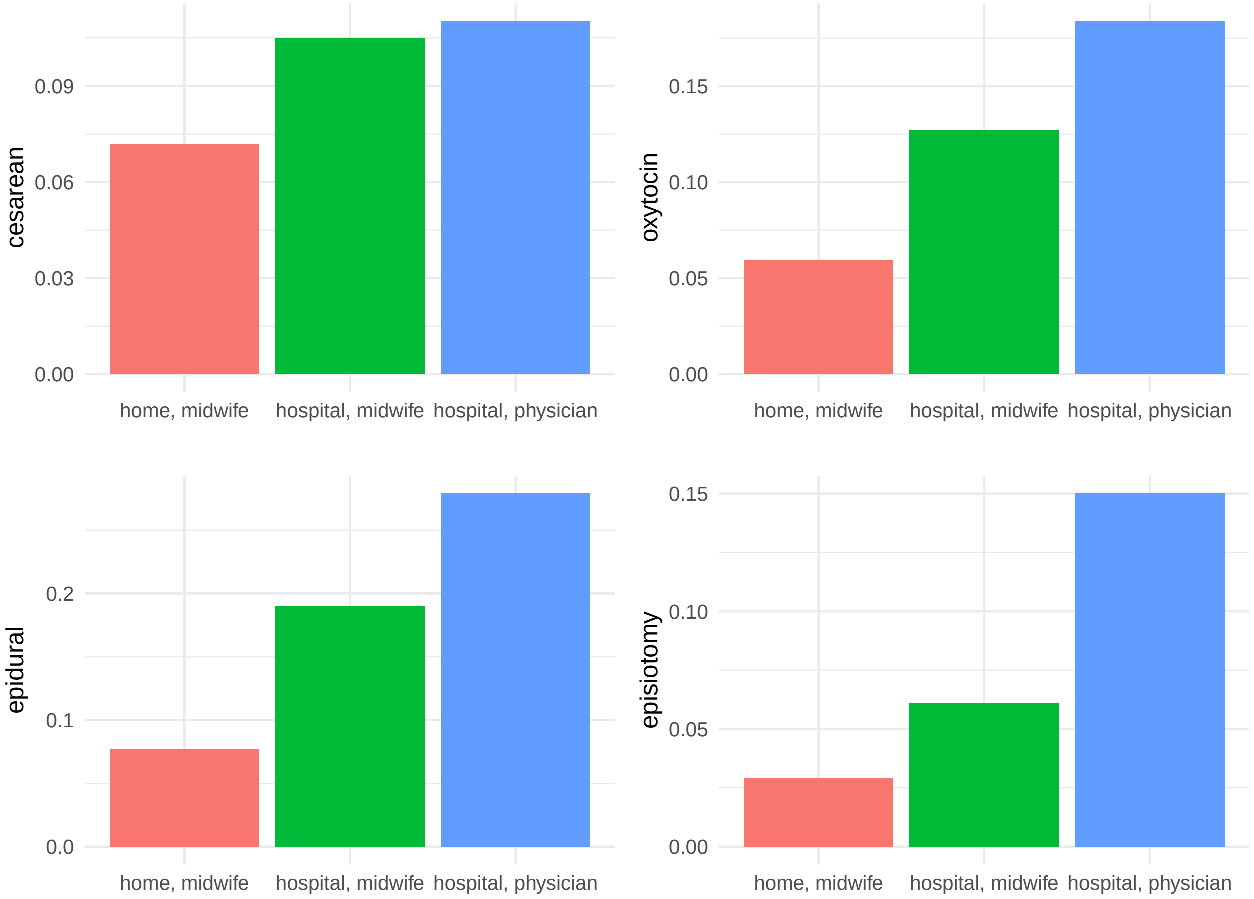
```
i1 <- ggplot(interventions_ratios) +
  geom_col(aes(birth_plan, cesarean, fill = birth_plan)) +
  theme(axis.text = element_text(size = 1)) +
  theme_minimal() +
  guides(fill="none") +
  xlab("")

i2 <- ggplot(interventions_ratios) +
  geom_col(aes(birth_plan, oxytocin, fill = birth_plan)) +
  theme(axis.text = element_text(size = 1)) +
  theme_minimal() +
  guides(fill="none") +
  xlab("")

i3 <- ggplot(interventions_ratios) +
  geom_col(aes(birth_plan, epidural, fill = birth_plan)) +
  theme(axis.text = element_text(size = 1)) +
  theme_minimal() +
  guides(fill="none") +
  xlab("")

i4 <- ggplot(interventions_ratios) +
  geom_col(aes(birth_plan, episiotomy, fill = birth_plan)) +
  theme(axis.text = element_text(size = 1)) +
  theme_minimal() +
  guides(fill="none") +
  xlab("")
```

```
(i1 + i2) / (i3 + i4)
```


cesarean
0.09
0.06
0.03
0.00
home, midwife
hospital, midwife
hospital, physician
oxytocin
0.15
0.10
0.05
0.00
home, midwife
hospital, midwife
hospital, physician
epidural
0.2
0.1
0.0
home, midwife
hospital, midwife
hospital, physician
episiotomy
0.15
0.10
0.05
0.00
home, midwife
hospital, midwife
hospital, physician