# Lab 3.A

## Workflow Development and EDA

Komal Bhosle

### RStudio Link

https://posit.cloud/spaces/603138/content/9848772

### Library Calls

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to beco
```

```
library(readxl)

library(openintro)
```

```
Loading required package: airports
Loading required package: cherryblossom
Loading required package: usdata
```

```
library(dplyr)
```

## Clean Variable Names

### The `rename()` function

```
airquality_rename <-
  airquality |>
  as_tibble() |>
  rename(
    ozone   = Ozone,
    solar.R = Solar.R,
    wind    = Wind,
    temp    = Temp,
    month   = Month,
    day     = Day
  )
```

### Inspecting Variable Names

```
colnames(airquality)
```

```
[1] "Ozone"   "Solar.R" "Wind"    "Temp"    "Month"   "Day"
```

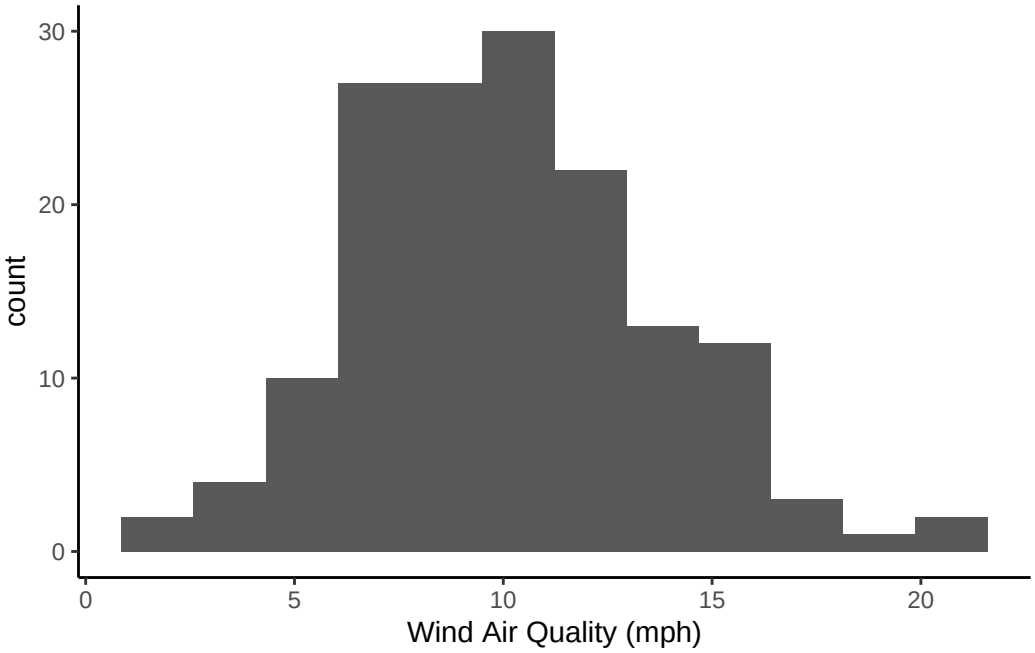
```
colnames(airquality_rename)
```

```
[1] "ozone"   "solar.R" "wind"    "temp"    "month"   "day"
```

### The `labs()` function

```
airquality_rename |>
  as_tibble() |>
  ggplot(
    mapping = aes(
      x = wind
```

```
    )
  ) +
  geom_histogram(
    bins = floor(sqrt(length(airquality_rename$wind)))
  ) +
  labs(
    x = "Wind Air Quality (mph)"
  ) +
  theme_classic()
```




## Clean Variable Datatypes

### Variable Coercion

```
table1_clean <-
  table1 |>
  mutate(
    country    = as_factor(country),
    year       = as.integer(year),
    cases      = as.integer(cases),
```

```
    population = as.integer(population)
  )
```

### Inspecting Variable Coercion

```
str(table1)
```

```
tibble [6 x 4] (S3: tbl_df/tbl/data.frame)
 $ country   : chr [1:6] "Afghanistan" "Afghanistan" "Brazil" "Brazil" ...
 $ year      : num [1:6] 1999 2000 1999 2000 1999 ...
 $ cases     : num [1:6] 745 2666 37737 80488 212258 ...
 $ population: num [1:6] 2.00e+07 2.06e+07 1.72e+08 1.75e+08 1.27e+09 ...
```

```
str(table1_clean)
```

```
tibble [6 x 4] (S3: tbl_df/tbl/data.frame)
 $ country   : Factor w/ 3 levels "Afghanistan",..: 1 1 2 2 3 3
 $ year      : int [1:6] 1999 2000 1999 2000 1999 2000
 $ cases     : int [1:6] 745 2666 37737 80488 212258 213766
 $ population: int [1:6] 19987071 20595360 172006362 174504898 1272915272 1280428583
```

```
is.ordered(table1_clean$country)
```

```
[1] FALSE
```

### Recoding and Decoding

```
diabetes2_clean <-
  diabetes2 |>
  mutate(
    treatment = case_when(
      treatment == "lifestyle" ~ "Lifestyle",
      treatment == "met" ~ "Met",
      treatment == "rosi" ~ "Rosi"
    ) |>
      as_factor(),
```

```
    outcome = case_when(
      outcome == "success" ~ "Success",
      outcome == "failure" ~ "Failure"
    ) |>
      as_factor()
  )
```

### Inspecting Categorical Variable Values

```
summary(diabetes2)
```

```
     treatment      outcome
 lifestyle:234   failure:319
 met       :232   success:380
 rosi      :233
```

```
summary(diabetes2_clean)
```

```
     treatment      outcome
 Met       :232   Success:380
 Rosi      :233   Failure:319
 Lifestyle:234
```

```
is.ordered(diabetes2_clean$treatment)
```

```
[1] FALSE
```

```
is.ordered(diabetes2_clean$outcome)
```

```
[1] FALSE
```

### Handling Missing Values

```r
airquality_pairwise_deletion <-
  airquality |>
  as_tibble() |>
  drop_na(Ozone)

airquality_colwise_deletion <-
  airquality |>
  as_tibble() |>
  drop_na(Ozone)
```

### Inspecting Missing Values

```r
anyNA(airquality)
```

```
[1] TRUE
```

```r
anyNA(airquality_pairwise_deletion)
```

```
[1] TRUE
```

```r
anyNA(airquality_colwise_deletion)
```

```
[1] TRUE
```

```r
summary(airquality)
```

```
     Ozone           Solar.R           Wind             Temp      
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00  
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00  
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00  
 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88  
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00  
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00  
 NA's   :37       NA's   :7                                       
     Month            Day       
 Min.   :5.000   Min.   : 1.0  
 1st Qu.:6.000   1st Qu.: 8.0  
```

```
 Median :7.000   Median :16.0
 Mean   :6.993   Mean   :15.8
 3rd Qu.:8.000   3rd Qu.:23.0
 Max.   :9.000   Max.   :31.0
```

```
summary(airquality_pairwise_deletion)
```

```
     Ozone           Solar.R          Wind             Temp
 Min.   :  1.00   Min.   :  7.0   Min.   : 2.300   Min.   :57.00
 1st Qu.: 18.00   1st Qu.:113.5   1st Qu.: 7.400   1st Qu.:71.00
 Median : 31.50   Median :207.0   Median : 9.700   Median :79.00
 Mean   : 42.13   Mean   :184.8   Mean   : 9.862   Mean   :77.87
 3rd Qu.: 63.25   3rd Qu.:255.5   3rd Qu.:11.500   3rd Qu.:85.00
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
                  NA's   :5
     Month            Day
 Min.   :5.000   Min.   : 1.00
 1st Qu.:6.000   1st Qu.: 8.00
 Median :7.000   Median :16.00
 Mean   :7.198   Mean   :15.53
 3rd Qu.:8.250   3rd Qu.:22.00
 Max.   :9.000   Max.   :31.00
```

```
summary(airquality_colwise_deletion)
```

```
     Ozone           Solar.R          Wind             Temp
 Min.   :  1.00   Min.   :  7.0   Min.   : 2.300   Min.   :57.00
 1st Qu.: 18.00   1st Qu.:113.5   1st Qu.: 7.400   1st Qu.:71.00
 Median : 31.50   Median :207.0   Median : 9.700   Median :79.00
 Mean   : 42.13   Mean   :184.8   Mean   : 9.862   Mean   :77.87
 3rd Qu.: 63.25   3rd Qu.:255.5   3rd Qu.:11.500   3rd Qu.:85.00
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
                  NA's   :5
     Month            Day
 Min.   :5.000   Min.   : 1.00
 1st Qu.:6.000   1st Qu.: 8.00
 Median :7.000   Median :16.00
 Mean   :7.198   Mean   :15.53
 3rd Qu.:8.250   3rd Qu.:22.00
```

```
 Max.   :9.000   Max.   :31.00
```

## Distinct Cases

### Preparation Code for Handling Duplicate Rows

```
airquality_with_duplicates <-
    bind_rows(
        airquality,
        slice_sample(
            .data = airquality,
            n     = 10
        )
    )
```

### Distinct Case Analysis

```
airquality_distinct <-
  airquality_with_duplicates |>
  distinct()
```

### Inspecting Distinct Cases

```
n_distinct(airquality_with_duplicates) < nrow(airquality_with_duplicates)
```

```
[1] TRUE
```

```
n_distinct(airquality_distinct) < nrow(airquality_distinct)
```

```
[1] FALSE
```

## Simple Wrangling Workflow

### Inspecting via `excel_sheets()`

```
excel_sheets("API_SE.ADT.LITR.ZS_DS2_en_excel_v2_293628.xls")
```

```
[1] "Data"                  "Metadata - Countries"  "Metadata - Indicators"
```

### Importing

```
wb_data <- read_excel("API_SE.ADT.LITR.ZS_DS2_en_excel_v2_293628.xls", skip = 2)

wb_data
```

```
# A tibble: 266 x 68
   `Country Name` `Country Code` `Indicator Name` `Indicator Code` `1960` `1961`
   <chr>          <chr>          <chr>            <chr>            <lgl>  <lgl>
 1 Aruba          ABW            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 2 Africa Easter~ AFE            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 3 Afghanistan    AFG            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 4 Africa Wester~ AFW            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 5 Angola         AGO            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 6 Albania        ALB            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 7 Andorra        AND            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 8 Arab World     ARB            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 9 United Arab E~ ARE            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
10 Argentina      ARG            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
# i 256 more rows
# i 62 more variables: `1962` <lgl>, `1963` <lgl>, `1964` <lgl>, `1965` <lgl>,
#   `1966` <lgl>, `1967` <lgl>, `1968` <lgl>, `1969` <lgl>, `1970` <dbl>,
#   `1971` <lgl>, `1972` <dbl>, `1973` <dbl>, `1974` <dbl>, `1975` <dbl>,
#   `1976` <dbl>, `1977` <dbl>, `1978` <dbl>, `1979` <dbl>, `1980` <dbl>,
#   `1981` <dbl>, `1982` <dbl>, `1983` <dbl>, `1984` <dbl>, `1985` <dbl>,
#   `1986` <dbl>, `1987` <dbl>, `1988` <dbl>, `1989` <dbl>, `1990` <dbl>, ...
```

```
wb_metadata_countries <- read_excel("API_SE.ADT.LITR.ZS_DS2_en_excel_v2_293628.xls", skip = 3

wb_metadata_countries
```

```
# A tibble: 266 x 68
   `Country Name` `Country Code` `Indicator Name` `Indicator Code` `1960` `1961`
   <chr>          <chr>          <chr>            <chr>            <lgl>  <lgl>
 1 Aruba          ABW            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 2 Africa Easter~ AFE            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 3 Afghanistan    AFG            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 4 Africa Wester~ AFW            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 5 Angola         AGO            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 6 Albania        ALB            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 7 Andorra        AND            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 8 Arab World     ARB            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 9 United Arab E~ ARE            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
10 Argentina      ARG            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
# i 256 more rows
# i 62 more variables: `1962` <lgl>, `1963` <lgl>, `1964` <lgl>, `1965` <lgl>,
#   `1966` <lgl>, `1967` <lgl>, `1968` <lgl>, `1969` <lgl>, `1970` <dbl>,
#   `1971` <lgl>, `1972` <dbl>, `1973` <dbl>, `1974` <dbl>, `1975` <dbl>,
#   `1976` <dbl>, `1977` <dbl>, `1978` <dbl>, `1979` <dbl>, `1980` <dbl>,
#   `1981` <dbl>, `1982` <dbl>, `1983` <dbl>, `1984` <dbl>, `1985` <dbl>,
#   `1986` <dbl>, `1987` <dbl>, `1988` <dbl>, `1989` <dbl>, `1990` <dbl>, ...
```

```
wb_metadata_indicators <- read_excel("API_SE.ADT.LITR.ZS_DS2_en_excel_v2_293628.xls", skip =

wb_metadata_indicators
```

```
# A tibble: 266 x 68
   `Country Name` `Country Code` `Indicator Name` `Indicator Code` `1960` `1961`
   <chr>          <chr>          <chr>            <chr>            <lgl>  <lgl>
 1 Aruba          ABW            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 2 Africa Easter~ AFE            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 3 Afghanistan    AFG            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 4 Africa Wester~ AFW            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 5 Angola         AGO            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 6 Albania        ALB            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 7 Andorra        AND            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 8 Arab World     ARB            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
 9 United Arab E~ ARE            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
10 Argentina      ARG            Literacy rate, ~ SE.ADT.LITR.ZS   NA     NA
# i 256 more rows
# i 62 more variables: `1962` <lgl>, `1963` <lgl>, `1964` <lgl>, `1965` <lgl>,
#   `1966` <lgl>, `1967` <lgl>, `1968` <lgl>, `1969` <lgl>, `1970` <dbl>,
#   `1971` <lgl>, `1972` <dbl>, `1973` <dbl>, `1974` <dbl>, `1975` <dbl>,
```

```
#   `1976` <dbl>, `1977` <dbl>, `1978` <dbl>, `1979` <dbl>, `1980` <dbl>,
#   `1981` <dbl>, `1982` <dbl>, `1983` <dbl>, `1984` <dbl>, `1985` <dbl>,
#   `1986` <dbl>, `1987` <dbl>, `1988` <dbl>, `1989` <dbl>, `1990` <dbl>, ...
```

### Wrangling

```
wb_data_clean <-
  wb_data |>
  pivot_longer(
    cols      = -c("Country Name", "Country Code", "Indicator Name", "Indicator Code"),
    names_to  = "year",
    values_to = "lit_rate"
  ) |>
  rename(country        = "Country Name",
         country_code   = "Country Code",
         indicator_name = "Indicator Name",
         indicator_code = "Indicator Code") |>
  mutate(
    country        = as_factor(country),
    country_code   = as_factor(country_code),
    indicator_name = as_factor(indicator_name),
    indicator_code = as_factor(indicator_code),
    year           = as.integer(year),
    lit_rate       = as.numeric(lit_rate)
  )
```

### Inspecting Wrangled Data

```
str(wb_data_clean)
```

```
tibble [17,024 x 6] (S3: tbl_df/tbl/data.frame)
 $ country       : Factor w/ 266 levels "Aruba","Africa Eastern and Southern",..: 1 1 1 1 1 1
 $ country_code  : Factor w/ 266 levels "ABW","AFE","AFG",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ indicator_name: Factor w/ 1 level "Literacy rate, adult total (% of people ages 15 and abo
 $ indicator_code: Factor w/ 1 level "SE.ADT.LITR.ZS": 1 1 1 1 1 1 1 1 1 1 ...
 $ year          : int [1:17024] 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 ...
 $ lit_rate      : num [1:17024] NA NA NA NA NA NA NA NA NA NA ...
```

```r
n_distinct(wb_data_clean) == nrow(wb_data_clean)
```

```
[1] TRUE
```

### DataViz Preparation

```r
wb_data_same_year <-
  wb_data_clean |>
  filter(year >= 1960 & year <= 2023)

country_names_clean<- unique(wb_data_clean$country)

wb_data_same_country <-
  wb_data_clean |>
  filter(country %in% country_names_clean)
```

### Inspecting DataViz Preparation

```r
unique(wb_data_clean$year)
```

```
 [1] 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974
[16] 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989
[31] 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004
[46] 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
[61] 2020 2021 2022 2023
```

```r
unique(wb_data_clean$country)
```

```
  [1] Aruba
  [2] Africa Eastern and Southern
  [3] Afghanistan
  [4] Africa Western and Central
  [5] Angola
  [6] Albania
  [7] Andorra
  [8] Arab World
```

```
 [9] United Arab Emirates
[10] Argentina
[11] Armenia
[12] American Samoa
[13] Antigua and Barbuda
[14] Australia
[15] Austria
[16] Azerbaijan
[17] Burundi
[18] Belgium
[19] Benin
[20] Burkina Faso
[21] Bangladesh
[22] Bulgaria
[23] Bahrain
[24] Bahamas, The
[25] Bosnia and Herzegovina
[26] Belarus
[27] Belize
[28] Bermuda
[29] Bolivia
[30] Brazil
[31] Barbados
[32] Brunei Darussalam
[33] Bhutan
[34] Botswana
[35] Central African Republic
[36] Canada
[37] Central Europe and the Baltics
[38] Switzerland
[39] Channel Islands
[40] Chile
[41] China
[42] Cote d'Ivoire
[43] Cameroon
[44] Congo, Dem. Rep.
[45] Congo, Rep.
[46] Colombia
[47] Comoros
[48] Cabo Verde
[49] Costa Rica
[50] Caribbean small states
[51] Cuba
```

```
[52] Curacao
[53] Cayman Islands
[54] Cyprus
[55] Czechia
[56] Germany
[57] Djibouti
[58] Dominica
[59] Denmark
[60] Dominican Republic
[61] Algeria
[62] East Asia & Pacific (excluding high income)
[63] Early-demographic dividend
[64] East Asia & Pacific
[65] Europe & Central Asia (excluding high income)
[66] Europe & Central Asia
[67] Ecuador
[68] Egypt, Arab Rep.
[69] Euro area
[70] Eritrea
[71] Spain
[72] Estonia
[73] Ethiopia
[74] European Union
[75] Fragile and conflict affected situations
[76] Finland
[77] Fiji
[78] France
[79] Faroe Islands
[80] Micronesia, Fed. Sts.
[81] Gabon
[82] United Kingdom
[83] Georgia
[84] Ghana
[85] Gibraltar
[86] Guinea
[87] Gambia, The
[88] Guinea-Bissau
[89] Equatorial Guinea
[90] Greece
[91] Grenada
[92] Greenland
[93] Guatemala
[94] Guam
```

```
 [95] Guyana
 [96] High income
 [97] Hong Kong SAR, China
 [98] Honduras
 [99] Heavily indebted poor countries (HIPC)
[100] Croatia
[101] Haiti
[102] Hungary
[103] IBRD only
[104] IDA & IBRD total
[105] IDA total
[106] IDA blend
[107] Indonesia
[108] IDA only
[109] Isle of Man
[110] India
[111] Not classified
[112] Ireland
[113] Iran, Islamic Rep.
[114] Iraq
[115] Iceland
[116] Israel
[117] Italy
[118] Jamaica
[119] Jordan
[120] Japan
[121] Kazakhstan
[122] Kenya
[123] Kyrgyz Republic
[124] Cambodia
[125] Kiribati
[126] St. Kitts and Nevis
[127] Korea, Rep.
[128] Kuwait
[129] Latin America & Caribbean (excluding high income)
[130] Lao PDR
[131] Lebanon
[132] Liberia
[133] Libya
[134] St. Lucia
[135] Latin America & Caribbean
[136] Least developed countries: UN classification
[137] Low income
```

```
[138] Liechtenstein
[139] Sri Lanka
[140] Lower middle income
[141] Low & middle income
[142] Lesotho
[143] Late-demographic dividend
[144] Lithuania
[145] Luxembourg
[146] Latvia
[147] Macao SAR, China
[148] St. Martin (French part)
[149] Morocco
[150] Monaco
[151] Moldova
[152] Madagascar
[153] Maldives
[154] Middle East & North Africa
[155] Mexico
[156] Marshall Islands
[157] Middle income
[158] North Macedonia
[159] Mali
[160] Malta
[161] Myanmar
[162] Middle East & North Africa (excluding high income)
[163] Montenegro
[164] Mongolia
[165] Northern Mariana Islands
[166] Mozambique
[167] Mauritania
[168] Mauritius
[169] Malawi
[170] Malaysia
[171] North America
[172] Namibia
[173] New Caledonia
[174] Niger
[175] Nigeria
[176] Nicaragua
[177] Netherlands
[178] Norway
[179] Nepal
[180] Nauru
```

```
[181] New Zealand
[182] OECD members
[183] Oman
[184] Other small states
[185] Pakistan
[186] Panama
[187] Peru
[188] Philippines
[189] Palau
[190] Papua New Guinea
[191] Poland
[192] Pre-demographic dividend
[193] Puerto Rico
[194] Korea, Dem. People's Rep.
[195] Portugal
[196] Paraguay
[197] West Bank and Gaza
[198] Pacific island small states
[199] Post-demographic dividend
[200] French Polynesia
[201] Qatar
[202] Romania
[203] Russian Federation
[204] Rwanda
[205] South Asia
[206] Saudi Arabia
[207] Sudan
[208] Senegal
[209] Singapore
[210] Solomon Islands
[211] Sierra Leone
[212] El Salvador
[213] San Marino
[214] Somalia
[215] Serbia
[216] Sub-Saharan Africa (excluding high income)
[217] South Sudan
[218] Sub-Saharan Africa
[219] Small states
[220] Sao Tome and Principe
[221] Suriname
[222] Slovak Republic
[223] Slovenia
```

[224] Sweden
[225] Eswatini
[226] Sint Maarten (Dutch part)
[227] Seychelles
[228] Syrian Arab Republic
[229] Turks and Caicos Islands
[230] Chad
[231] East Asia & Pacific (IDA & IBRD countries)
[232] Europe & Central Asia (IDA & IBRD countries)
[233] Togo
[234] Thailand
[235] Tajikistan
[236] Turkmenistan
[237] Latin America & the Caribbean (IDA & IBRD countries)
[238] Timor-Leste
[239] Middle East & North Africa (IDA & IBRD countries)
[240] Tonga
[241] South Asia (IDA & IBRD)
[242] Sub-Saharan Africa (IDA & IBRD countries)
[243] Trinidad and Tobago
[244] Tunisia
[245] Turkiye
[246] Tuvalu
[247] Tanzania
[248] Uganda
[249] Ukraine
[250] Upper middle income
[251] Uruguay
[252] United States
[253] Uzbekistan
[254] St. Vincent and the Grenadines
[255] Venezuela, RB
[256] British Virgin Islands
[257] Virgin Islands (U.S.)
[258] Viet Nam
[259] Vanuatu
[260] World
[261] Samoa
[262] Kosovo
[263] Yemen, Rep.
[264] South Africa
[265] Zambia
[266] Zimbabwe

```
266 Levels: Aruba Africa Eastern and Southern ... Zimbabwe
```

```
unique(wb_data_same_year$year)
```

```
 [1] 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974
[16] 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989
[31] 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004
[46] 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
[61] 2020 2021 2022 2023
```

```
unique(wb_data_same_country)
```

```
# A tibble: 17,024 x 6
   country country_code indicator_name             indicator_code  year lit_rate
   <fct>   <fct>        <fct>                      <fct>          <int>    <dbl>
 1 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1960       NA
 2 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1961       NA
 3 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1962       NA
 4 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1963       NA
 5 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1964       NA
 6 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1965       NA
 7 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1966       NA
 8 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1967       NA
 9 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1968       NA
10 Aruba   ABW          Literacy rate, adult tota~ SE.ADT.LITR.ZS  1969       NA
# i 17,014 more rows
```

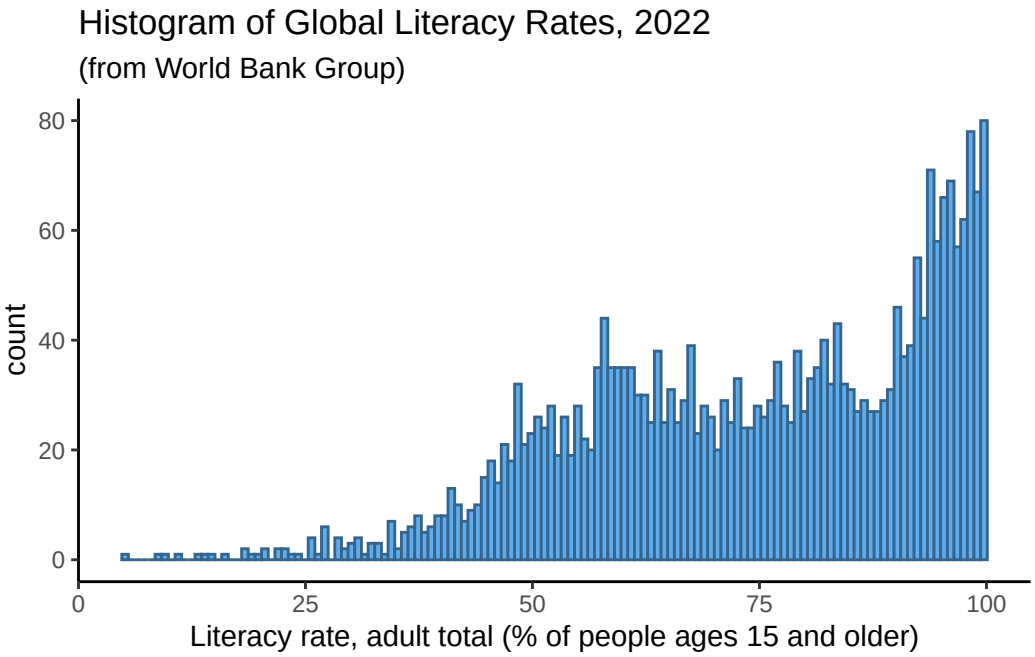
## Simple DataViz Workflows

### Univariate Plot – Histogram

```
wb_data_same_year |>
  as_tibble() |>
  ggplot(
    mapping = aes(
      x = lit_rate
      )
    ) +
  geom_histogram(
```

```
    bins  = floor(sqrt(length(wb_data_same_year$lit_rate))),
    fill  = "steelblue2",
    color = "steelblue4"
  ) +
  labs(
    x        = "Literacy rate, adult total (% of people ages 15 and older)",
    title    = "Histogram of Global Literacy Rates, 2022",
    subtitle = "(from World Bank Group)"
  ) +
  theme_classic()
```

```
Warning: Removed 14289 rows containing non-finite outside the scale range
(`stat_bin()`).
```



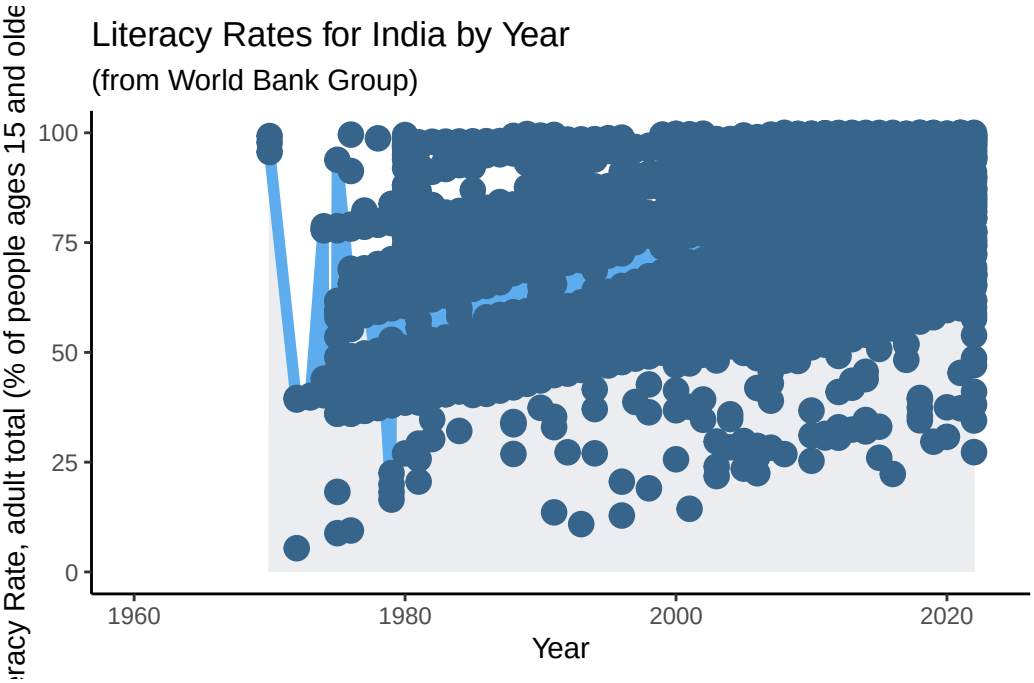

**Bivariate Plot – Line Plot**

```
wb_data_same_country |>
  as_tibble() |>
  ggplot(
    mapping = aes(
```

```
      x     = year,
      y     = lit_rate
      )
    ) +
  geom_area(
    fill      = "#ebedf0",
    color     = "steelblue2",
    linewidth = 2
  ) +
  geom_point(
    color = "steelblue4",
    size  = 4
  ) +
  labs(
    x        = "Year",
    y        = "Literacy Rate, adult total (% of people ages 15 and older)",
    title    = "Literacy Rates for India by Year",
    subtitle = "(from World Bank Group)"
  ) +
  theme_classic()
```

```
Warning: Removed 14289 rows containing non-finite outside the scale range
(`stat_align()`).
```

```
Warning: Removed 14289 rows containing missing values or values outside the scale range
(`geom_point()`).
```


Literacy Rates for India by Year
(from World Bank Group)
racy Rate, adult total (% of people ages 15 and olde
100
75
50
25
0
1960
1980
2000
2020
Year