Foundations of Public Health (W1)

Identify the main areas of public health today and from a historical context

Public Health

- Doing something to prevent people from getting sick
- Improve the health of an entire population
- Reduce health inequalities in a population
- Step beyond the individual-level focus of mainstream medicine
- Important tools are Epidemiology and Biostatistics
  - **Epidemiology** is: the study of the distribution and determinants of health and disease in populations, and the application of this study to control health problems

The birth of Modern Epidemiology

- John Snow and the Broad Street pump

Differentiate the concepts of cause and association

Critical issues in Epidemiology: Causation

- The concept of causation in epidemiology is critical
- There is actually not one single cause for any disease

Causation-Sufficient and Necessary

- **Causation**: Something that either **alone** or **in combination** with other things produce an outcome. Important for prevention, correct diagnosis and treatment
- **Association**: Statistical relationship between two or more events. (Lung cancer is highly associated with cigarettes)
- A risk factor is **sufficient**, if the presence of this factor **alone** is enough to **result** in the **disease**
- A risk factor is **necessary** if the disease is **never present when the factor is not present**
  - To get tuberculosis it is **necessary** to be exposed to Mycobacterium tuberculosis but the exposure in itself is **not sufficient** for the disease state to occur
- Cigarette smoking is **neither necessary nor sufficient** for the development of lung cancer
- A necessary and sufficient cause of Ebola fever is the Ebola virus
- A necessary and sufficient cause of Huntington's chorea is the genetic mutation that causes this condition (dominant)

As few diseases have a single cause, we tend to talk about "risk factors"

- Pre-disposing: age, sex
- Enabling (disabling): low income, poor nutrition
- Precipitating: exposure to a disease agent
- Reinforcing: repeated exposure

Causation-Establishing Evidence (Bradford Hill criteria)

- **Temporality**: Does the cause proceed the effect?
- **Plausibility**: Is the association consistent with existing knowledge?
- **Consistency**: Have similar results been shown in other studies
- **Strength of association**: What is the strength of the association between the cause and effect (degree to which the values of two variables vary or **change together**)
- **Dose response**: Does increased exposure = increased effect?
- **Reversibility**: Does removal of a cause decrease the risk of the effect?

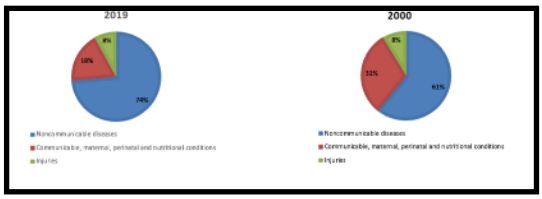
If there is an association between a possible cause and an effect

- If there is an association between a possible cause and an effect,
  - Could it be due to bias?
  - Could it be due to confounding?
  - Could it be the result of chance?
  - Is the relationship causal?

Describe how public health is quantified by identifying different data types

Public Health Domains and Indicators

1. Injury
2. Communicable diseases, maternal and perinatal conditions
3. Non-communicable disease



Deaths worldwide:

- Injuries result in:
  - An estimated 12,000 (or 8%) of deaths each year in Australia
  - 460,000 hospital admissions, annually
- Injuries are the principal cause of death in almost half of the people under 45 years of age;
  - High road toll (8.7 per 100,000)
  - High male suicide (19.4 per 100,000)
  - Only a small change in the incidence of children poisoning since early 1990s. (267 per 100,000)

Smoking - The Epidemic of the 20th Century

- 100 million people died of tobacco-related causes during the 20th century
- WHO predicts that without intervention tobacco will kill 1 billion this century
- 80% of those deaths will occur in developing countries

Epidemiology in your professional lifetime impact of climate change

- Direct (impact on coastal communities)
- Indirect (food, water, social, economic) effects
- Temperature change (heat waves, storms, floods)
- Nutrition and food security (crop yields)
- Water availability and quality
- Air quality (pollutants, aeroallergens)
- Vector, rodent and bird-borne diseases
- Exposure to UV radiation

Types of Data

Two types of data

1. **Categorical**
2. **Numerical**

- What type of graph?
- What type of summary statistics
- What type of statistical analysis

Categorical data subtype

- **Nominal** (order of categories **doesn't** matter)
  - eg. Blood group (A, B, O, AB)
  - eg. Diabetic (Yes/No) Also known as "binary"
- **Ordinal** (order of categories **does** matter)
  - eg. BMI (underweight, normal, overweight, obese)
  - Pain severity (None, Mild, Moderate, Severe)

Numerical data subtypes

- **Discrete** (counts)
  - eg. Number of hospital admissions
- **Continuous** (measurements)
  - eg. Weight, Height, BMI, blood pressure, cholesterol, etc

Data dictionary example

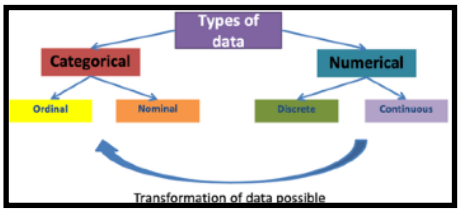
Variable name	Variable description
Patient ID	001, 002, 003, 004, etc
Gender	Male = 1, Female = 2
Weight (kg)	
Height (m)	
Body Mass Index (BMI)	
BMI category	Underweight (BMI < 18.5) = 0 Healthy (BMI ≥ 18.5 to BMI < 25) = 1 Overweight (BMI ≥ 25 to BMI < 30) = 2 Obese (BMI ≥ 30 to BMI < 40) = 3 Morbidly obese (BMI ≥ 40) = 4
Blood group	A = 1, B = 2, AB = 3, O = 4
Number of hospital admissions	0, 1, 2, 3, 4, ... etc
Cancer stage	I = 1, II = 2, III = 3, IV = 4
Received chemotherapy	Yes = 1, No = 2
Received radiotherapy	Yes = 1, No = 0

Categorical | Numerical

Data dictionary example

Variable name	Variable description	
Patient ID	001, 002, 003, 004, etc	
Gender	Male = 1, Female = 2	Binary
Weight (kg)		
Height (m)		
Body Mass Index (BMI)		
BMI category	Underweight (BMI < 18.5) = 0 Healthy (BMI ≥ 18.5 to BMI < 25) = 1 Overweight (BMI ≥ 25 to BMI < 30) = 2 Obese (BMI ≥ 30 to BMI < 40) = 3 Morbidly obese (BMI ≥ 40) = 4	
Blood group	A = 1, B = 2, AB = 3, O = 4	
Number of hospital admissions	0, 1, 2, 3, 4, ... etc	
Cancer stage	I = 1, II = 2, III = 3, IV = 4	
Received chemotherapy	Yes = 1, No = 2	Binary
Received radiotherapy	Yes = 1, No = 0	Binary

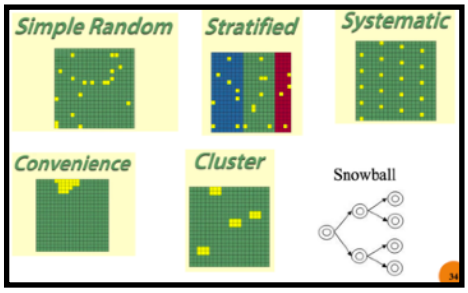
Categorical Nominal | Categorical Ordinal | Numerical Discrete | Numerical Continuous



Data transformation - eg. Body Mass Index
- BMI (**continuous**)
  - Calculated by: Weight (kg) / [Height (m)]^2
- BMI categories (**ordinal**)
  - Underweight, Healthy weight, Overweight, Obese, Morbidly obese
- In which direction is transformation possible?
- In which direction is transformation not possible?
  - Transformation of **Continuous -> Categorical is POSSIBLE**
  - Transformation of **Categorical -> Continuous is NOT POSSIBLE**

Population vs Samples

Population and Samples
- The population: The entire collection of patients of interest, as defined in the study design
  - eg. All diabetics in Australia
- Actual population of interest often large
  - Can't measure everyone - expensive, time, etc
  - Exception = Census (every 5 years)!
- Solution: Take a **sample** that is representative of part of the population of interest. Collect data for this sample
  - Eg. Sample of 100 diabetics in Australia
  - Use this data to make inferences about the population of interest
- Don't forget: Behind statistics and studies are PEOPLE



Sampling level
- Sampling is normally at the individual level
- Other times (when investigating an intervention), it is at a "cluster" level
  - House
  - School classrooms, or schools
  - Hospital wards, or hospitals

Simple: Sampling frame is everyone doing 1042. Numbered 1 to 800. **Randomly** select 10 numbers
Stratified sample: Have a list of everyone in 10am workshop and 2pm workshop. Randomly select 5 students from each workshop using "simple" method
Systematic: Select **every 20th** student on the roll, starting with 15th
Convenient: Penny to choose a couple of groups of 6 students (front of row, near the aisle). Ask - why don't I choose students in centre middle?
Cluster: Workshop rooms had 15 tables, 6 students on each. **Select 2 tables**
Snowball sampling: Criteria = wear glasses and **recruit others** who also have glasses

Types of Samples - Snowball Sampling
- Find people who meet your criteria. Ask them to refer you to others
- eg. The Burnet Institute recruit people for many studies where there is no "list: of eligible people
- For example. Injecting Drug Users (IDU), Men who have Sex with Men (MSM)
- Think about how viral social media posts work amongst your friends

Understanding terms
- **Population**: complete collection of people of interest
- **Sample**: selection of participants that are taken from a population of interest
- **Data**: **pieces of information** about individuals which are organised into variables
- **Dataset**: a **set** of data
- **Variable**: anything that is **measured** or recorded on an individual
- **Derived variable**: a variable that is **created** from **another**. eg. **BMI categories from BMI continuous**
- **Observation**: the value, for a particular, of a particular variable
- **Outcome variable**: the **focus** of the analysis

- **Exposure variable**: the factors that may **influence** the outcome
- **Numerical data**: a measurement with **numbers** in it
- **Categorical**: a **non-numerical** variable; has groups

Describing Population Health (W2)

Pre-tutorial Video:

Categorical data - Relative **Frequency**
- Tabulate the number of observations in each category
- Relative frequency = percentage in each category

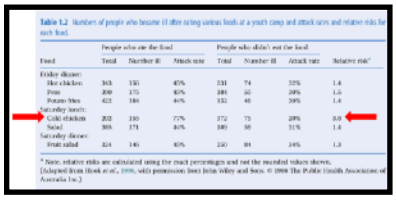
Table 3.1 Method of delivery of 600 babies born in a hospital. Source: *Essential Medical Statistics*, p. 16

Method of delivery	No. of births	Percentage
Normal	478	79.7%
Forceps	65	10.8%
Caesarean section	57	9.5%
Total	**600**	**100.0%**

Categorical data - **crosstable**
- Tabulation of the number of observations - in each combination of categories. eg. Disease and pesticide exposure
- Can calculate:
  - row % (% developed disease, by pesticide exposure status)
  - Column % (% exposed to pesticides, by diseases status)

Table 5.9 The results of a hypothetical study of the effects of pesticide exposure.

	Developed disease	Did not develop disease	Total
Exposed to pesticides	40	2,460	2,500
Not exposed	60	7,440	7,500
Total	100	9,900	10,000

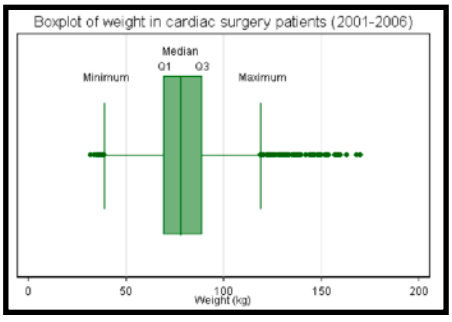


Categorical data - expanded table example
- Many different foods summarised in table:
  - Attack rate (% of those who ate the food that got ill)
  - Relative risk (Risk exposed / Risk unexposed)

Numerical data - Descriptive statistics for continuous data
- **Central tendency**
  - Mean (Average = Sum observations/No. Obs)
    - Suitable when data normal, not when skewed
  - Median (middle value) - good for skewed data/outliers
  - Mode (most frequent measurement [or range])
- **Variation**
  - Standard deviation. Variation in observations from mean
  - Interquartile range [IQR] (3rd quartile - 1st quartile)

Importance of data checking
- Unusual values and missing data
- ASCTS (cardiac surgery) database data from 2001-2004
- Maximum Body Mass Index = 2844.44 kg/m^2
- Many values considered “outliers”
  - 42 BMI values >=48
  - 13 BMI values >= 100
- Requested all BMI > 48 kg/m^2 to be double checked
  - Some height was swapped with weight entries
  - Maximum BMI due to height = 15cm recorded
  - After corrections, statistics more believable

Types of graphs
- Each type of graph has a different purpose
- Each graph displays data differently
- Some suitable for categorical data, other numerical
- Some common graphs include:
  - **Bar chart** (single or multiple)
  - **Pie chart**
  - **Scatterplot**
  - **Histogram**
  - **Box plot**
    - Q1 = 1st quartile (25th Percentile)
    - Q2 = 2nd quartile Median value
    - Q3 = 3rd quartile (75th percentile)
    - Dots represent outliers



Define and describe descriptive epidemiological concepts and terms

Descriptive **epidemiology**

- Aim:
  - Information about **disease patterns** - characteristics of persons, place, time and using descriptive statistics
  - To answer the question, what is the health situation

Descriptive **Biostats**

- Aim:
  - Identify **healthy objectives** i.e. National Health Priority Areas
  - To assess the health status of a population I.e. Millennium Development Goals
  - Allows **allocation of resources** efficiently by targeting populations

Prevention Approach

- The first stage of epidemiological investigation. It focuses on describing disease distribution by characteristics relating to **time, place, and person**
  - Who? What? Where? When?
- It is called **Descriptive Epidemiology**

Population at Risk

- Population at risk:
  - Should only include people who are potentially **susceptible** to the disease being studied
  - Can be defined by demographic, geographic or environmental factors
- E.g. Occupational injures = workforce

Ability to distinguish between mortality and morbidity

Mortality

- **Definition**: Death or state being **subject to death**
- How we use it:
  - The number of people who **died** within a population
  - Common measure
  - Uninformative for some diseases
    - eg. Osteoarthritis or psoriasis
    - Does not mirror underlying incidence of disease
      - What happens if a more effective treatment is introduced
      - Lag behind changes in incidence
- Crude
- Not always complete

Many different death rates

- **Crude** death rate = number of deaths / number of persons at risk of dying
- **Age-specific** death rate - death rate for specific group of population
- **Proportionate mortality** - number of deaths from a given cause, per 100 or 1000 **total** deaths
- **Infant mortality** - rate of death in children during the first year of life
- **Child mortality** rate - deaths of children aged 1-4 years (Basic health indicator)
- **Maternal mortality** rate - risk of mothers dying from causes associated with **delivering babies**, complications of pregnancy or childbirth
- **Adult mortality** rate - probability of dying between the ages of 15 and 60 years, per 1000
- **Life expectancy** - average number of years an individual of given age is expected to live if current mortality rates continue
- **Age-standardised** mortality rates - summary measure of death rate that a population would have if it had a standard age structure

Public Health Domain: Injury

- In 2017-2018 there were:
  - 532,500 hospitalisations
  - 13,000 injury deaths
  - 54% of injury deaths were for people aged over 65

- Top 3 causes of hospitalisation were falls, suicide and transport accidents

Crude Mortality Rate (CMR)
- **Crude death rate** = **number of deaths / number of persons at risk of dying**
- Calculate the crude mortality rate (per 100,000) in Australia in 2005 from drowning, using the following information
- Drowning deaths: 232
- Mid-year population: 20,394,791
- Crude mortality rate = drowning deaths / mid-year population

Age-specific mortality rate
- Age-specific death rate - death rate for **specific group** of population
- Calculate the age-specific mortality rate (per 100,000) in Australia in 2005 from drowning, using the following information
- Age group 0-4 years: 23 deaths
- Mid-year population: 1,285,545
- Age-specific mortality rate = number of drowning deaths in age group/people in age group
- 23/1,285,545 * 100,000 = 1.79 per 100,000

Life Expectancy
- Life expectancy - average number of years an individual of given age is expected to live if current mortality rates continue

Describe and calculate incidence and prevalence measures

Morbidity
- **Definition**: Any departure, subjective or objective, from a state of physiological or psychological well being
- eg. Headache, back pain, tooth ache

Disability
- Definition: Any temporary or permanent **reduction** of a person's capacity to **function**
- eg. Stuck in bed with cold, broken leg

Morbidity
- Death rates are useful for high case-fatality
- Not so useful for low case-fatality chronic conditions
  - eg. Mental illness, rheumatoid arthritis
- Morbidity (burden of **diseases measure**) is more useful for **chronic conditions with low case fatality**
- Looks at impairment, disability and handicap
- In terms of morbidity, mental health issues are some of the most important causes of morbidity in the community

Common measures of morbidity and or disability
- Incidence
- Prevalence
- Hospital discharge
- Number of office visits/consultations
- Bed disability days
- Work loss days
- Restricted activity days

Case Study: Lead Poisoning
- 100% preventable
- There are NO safe levels
- Exposures:
  - Occupational (smelters)
  - Paint
  - Lead gasoline
  - Polluted soil

Incidence
- New cases of a disease or event or death
- Frequency of development of a new case or occurrence of disease in a population over time
- Express this as either:
  - Rate (Incidence rate)
  - Proportion (Cumulative Incidence)

Measuring Disease Frequency
- **Cumulative Incidence (Risk)** - proportion of new cases within a specific time period
  - Range of values 0 - 1, Assumes that the entire population at risk has been followed up for the entire study period
- **Incidence Rate** - 'speed' at which disease occurs
  - Range of values 0 - infinity. Assumes all people in population are at risk at the beginning of the period and remain at risk

$$\frac{\text{Number of new cases during a specified time period}}{\text{Number of people at risk (disease free) at start of the study time period}}$$

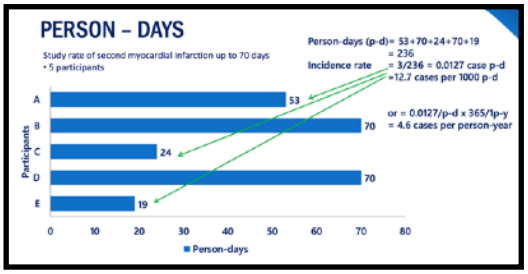
$$\frac{\text{Number of new cases}}{\text{Total Person-time of Observation/Risk}}$$

Cumulative Incidence
- Defintion:
  - Proportion of a population at risk of getting a disease that become diseased **over a period of time**
  - E.g. 4,000 children are observed for 2 years. Over the course of 2 years 1600 children get influenza
  - **Cumulative incidence** = 1600/4000 = 0.4 or 40% over a 2 year period

Incidence Rate
- The number of new cases of a disease divided by the **person-time** that the population at risk is observed
- Person-time = true rate
  - Estimate of the actual time-at-risk in years of all persons that contribute to the study
  - Example: 4,000 children are observed for 2 years contributing a total of **7,892 person-years**. Over the course of a year 1600 children get influenza
  - **Incidence Rate** = 1600/7892 = 0.203 cases per person-years
    - 20.3 cases per 100 person-years

- During 2007, 1.7 million people (0.22%/216 new cases per 100,000 people) in Sub-Saharan Africa were newly diagnosed as HIV positive. What is the:
  - Rate?
    - 216 new cases per 100,000 people per year
  - Cumulative incidence
    - 0.22%

Person-Time
- Is the actual time at risk that all participants contribute to a study
- For example:
  - 1 person followed for a 1 year without disease
  - 2 people of 6 months without disease
  - 6 people followed for 2 months without disease
  - 1 person for 2 months, 1 person for 7 months and 1 person for 3 months all without disease
- Why?
  - Instead of counting the actual number of people at risk we count the **length of time they were at risk**
  - More accurate measure of how quickly the disease is occurring as we can identify when **risk status changes** for each participant
  - Plus more flexibility on how we can interpret the results



Number of cases of a disease present in the population
Number of persons in a population

Prevalence
- Existing cases
- Frequency of an **existing** condition/disease at one point in time during a given period
- Factors that affect prevalence:
  - Disease occurrence e.g. asthma
  - Duration of illness

Relationship between prevalence and incidence
- If two diseases have the **same incidence**, but **one lasts 3 times longer**… at any point in time you are **more likely to find people suffering from the long-lasting disease**
- **Prevalence** = **Incidence rate** X **average duration of disease**
  - Only true if:
    - Stationary population
    - Prevalence is low
- The example: Hepatitis A versus Hepatitis C in the US (infectious liver disease)
- Hepatitis A (acute): vaccine preventable, good hygiene and sanitation
- Hepatitis C (chronic): treated with medication, no vaccine, harm reduction strategies: needle exchange programs

Describe the use of QALYs and DALYs as measures of morbidity

Measures of morbidity
- More comprehensive measure of sieges impact than mortality:
  - Potential Years of Life Lost (**PYLL**)
  - Quality-Adjusted Life Years (**QALYs**)
  - Disability-Adjusted Life Years (**DALYs**)
    - Years of Life Lost (**YLL**)
    - Years Lost to Disability (**YLD**)

PYLL = Total # of death from a specific cause in each age group
X
average # of years of life lost as a result of each of these deaths

Potential Years of Life Lost (**PYLL**)
- Measure of health expectancy or **number of years of potential life they lost** if they die before a certain age
  - Most reports use death **before 65** years premature deaths
- PYLL = **Total # of death** from a specific cause in each age group X average **# of years of life lost** as a result of each of these deaths
- An example: Fatal car accident at 35 years of age versus 65 years

Quality-Adjusted Life Years (**QALYs**)
- Weigh each year of life by the perceived **quality** of that life from a value of 1 for perfect down to 0
- Example, imagine a person aged 55 years who lives for 10 years after a stroke and dies at age 65
  - e.g. 2 years of perfect health is 2 QALYs

Disability-Adjusted Life Years (**DALY**)
- **Disability**-Adjusted Life Years is a measure of **overall disease burden,** expressed as the cumulative **number of years lost due** to **ill-health, disability or early death**
- **DALY** = YLD (**Years Lived with Disability**) + YLL (**Years of Life Lost**)
- Health gap indicator are increasingly used for **estimate of burden** of diseases so that we can target interventions
- Highlight the enormous **burden of ill-health** due to some common but fatal conditions that don't appear on our mortality rate indicators

Making Data Easy to Understand: Graphs and Tables

Injury In Australia

- 454,000 Australians were hospitalised due to injury in 2011-12
- 2 of the main causes of injury are falls and transport accidents

What data will be available

- **Mortality**
  - Deaths while cycling
- **Morbidity - burden of disease**
  - Serious injury
  - Moderate injury
  - Mild injury
  - No injury
  - "Near miss"
- Death, hospital, doctor, Physio (Etc.) records

Graphs and Statistical Presentation of Data

Frequency Tables

- Frequency Table and Relative frequency example:
- **One** variable
- Gender of students at Monash Uni from Nov 2015
- Type of variable = **CATEGORICAL -> NOMINAL**
  - NOT binary

$$= \frac{\text{Number of observations in category}}{\text{Total number of observations}} \times 100\%$$

Gender category	Number in category (n)	Calculation column	Relative frequency (%)
M = Male	64,207	$\frac{64207}{154121} \times 100\%$	41.7 %
F = Female	89,806	$\frac{89506}{154121} \times 100\%$	58.3 %
X = Indeterminate / Intersex / Unspecified	106	$\frac{106}{154121} \times 100\%$	0.069 %
U = Unknown	2	$\frac{2}{154121} \times 100\%$	0.001 %
Total	**154,121**		

			Respiratory symptoms in the past 12 months		
			Yes	No	Total
Gender of child	Female	n	81	254	335
		Row %	24.18 %	75.82 %	100.00 %
		Column %	55.86 %	51.73 %	52.67 %
		Total %	12.74 %	39.94 %	52.67 %
	Male	n	64	237	301
		Row %	21.26 %	78.74 %	100.00 %
		Column %	44.14 %	48.27 %	47.33 %
		Total %	10.06 %	37.26 %	47.33 %
	Total	n	145	491	636
		Row %	22.80 %	77.20 %	100.00 %
		Column %	100.00 %	100.00 %	100.00 %
		Total %	22.80 %	77.20 %	100.00 %

Cross Tabulation of two categorical variables

- Question: In the "Peru Lung" dataset from Essential Medical Statistics, is there a relationship between gender and respiratory symptoms
- Good for **two** variables
- What overall % are we interested in?
  - % gender (335 / 636 = 52.67%) —> column %
  - % respiratory symptoms (145 / 636 = 22.80%) —> row %
- Row % (respiratory symptoms %, then by gender)
  - Each row adds up to 100%
- Column % (gender%, then by respiratory symptoms)
  - Each column adds up to 100%

Which % is better? It depends on your question

- Overall, 22.80% (145/636) had **respiratory symptoms**
  - 77.20% (491/636) had **no respiratory symptoms**
  - 24.18% (81/335) of **females** had **respiratory symptoms**
  - 21.26% (64/301) of **males** had **respiratory symptoms**
- Overall, 52.67% (335/636) were female
  - 47.33% (301/636) were male
  - 55.86% of **those with respiratory symptoms** were **female**
  - 51.73% of **those without respiratory symptoms** were **female**
- Which overall % makes more sense
  - Respiratory symptoms
  - Therefore row % best in this case

Summary statistics

- Provides key information about the data
- Concise information