Habits Related to Health of students of M.H.Hall AMU, Aligarh

Mohd Mursaleen
03/10/2021

CHAPTER 1

INTRODUCTION

1.1 Health

Health, is defined by the World Health Organization (WHO), as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." This definition has been subject to controversy, as it may have limited value for implementation. Health may be defined as the ability to adapt and manage physical, mental and social challenges throughout life. Generally, the context in which an individual lives is of great importance for both his health status and quality of their life, It is increasingly recognized that health is maintained and improved not only through the advancement and application of health science but also through the efforts and intelligent lifestyle choices of the individual and society. According to the World Health Organization, the main determinants of health include the social and economic environment, the physical environment and the person's individual characteristics and behaviors. More specifically, key factors that have been found to influence whether people are healthy or unhealthy include the following:

- Income and social status
- Social support networks
- Education and literacy
- Employment/working conditions
- Social Environment
- Physical Environment • Personal health practices and coping skills
- Healthy child development
- Biology and genetics
- Health care services
- Gender
- Culture

1.2 Hygiene

Hygiene is a set of practices performed to preserve health. According to the World Health Organisation(WHO), "Hygiene refers to conditions and practices that help to maintain health and prevent the spread of diseases." Personal hygiene refers to maintaining the body's cleanliness. Many people equate hygiene with 'cleanliness,' but hygiene is a broad term. It includes such personal habit choices as how frequently to take a shower or bathe, wash hands, trim fingernails and change and wash clothes. It also includes attention to keeping surfaces in the home and workplace, including bathroom facilities, clean and pathogen-free. Some regular hygiene practices may be considered as a good habit by a society, while the neglect of hygiene can be considered disgusting, disrespectful, or threatening. Hygiene is a concept related to cleanliness, health and medicine. It is as well related to personal and professional care practices. In medicine and everyday life settings, hygiene practices are employed as preventative measures to reduce the incidence and spreading of disease. Hygiene practices vary, and what is considered acceptable in one culture might not be acceptable in another. In the manufacturing of food, pharmaceutical, cosmetic and other products, good hygiene is a critical component of quality assurance The terms cleanliness and hygiene are often used interchangeably, which can cause confusion. In general, hygiene refers to the practices that prevent spread of disease-causing organisms. Cleaning processes (e.g., hand-washing) remove infectious microbes as well as dirt and soil, and are thus often the means to achieve hygiene. Other uses of the term appear in phrases including body hygiene personal hygiene, sleep hygiene, mental hygiene,dental hygiene and occupational hygiene used in connection with public health. Hygiene is also the name of a branch of science that deals with the promotion and preservation of health .

2. Parameters of Health

- Body Mass Index
- Exercise
- Sleep
- Environment

2.1 BMI(Body Mass Index)

The Body Mass Index(BMI) is a physical measurement used to assess an individual's total amount of body fat. The BMI was invented by Belgian polymath Adolphe Quetelet in the 1800s, and consequently is sometimes known as the Quetelet Index[3]. The BMI is calculated by dividing your weight in kilograms (kg) by your height in metres squared (m2). It is expressed as kg/m2. BMIcut-off values are given in Table 1. What does your BMI mean? The BMI scores give an indirect measure of body fat. Depending on the BMI value calculated you may be underweight, healthy weight, overweight or obese. The cut off values are as follows: [4]

Table 1: BMI cut-off values

BMI Classification

Classification of BMI

Range	Category
< 18.5	Underweight
18.5-24.9	Healthy weight
25.0-29.9	Overweight
30.0-34.9	Obese class 1
35.0-39.9	Obese class 2

Range	Category
> 40.0	Obese class 3

Why are BMIs useful?

The BMI is a simple, inexpensive screening tool used to identify possible weight problems for both adults and children. A BMI measurement is useful to assess who needs further testing to identify health risks such as heart diseases.Individuals at risk will need further assessment. Assessments may include skin fold thickness test, diet, physical activity level, family history and other appropriate health screenings. Risks associated with extreme BMIs: Being either overweight (with a BMI of 25 or above) or underweight (with a BMI lower than 18.5) can affect your health. Overweight :Individuals who are over-weighted or obese have a greater risk of disease compared to those in the healthy weight range. The risk of disease increases with increasing BMI. Those classified overweight $(BMI = 25-29.9)$ may also be considered pre-obese and are at an increased risk of disease. In Category 1 obesity $(BMI = 30-34.9)$ there is a moderate risk of disease, which increases to severe and very severe risk at obesity stages 2 $(BMI = 35-39.9)$ and 3 $(BMI \geq 40)$ respectively. It is important to note that BMI does not determine risk alone. Other factors such as what one eats, how much they exercise and whether or not there is a history of disease in their family also influences an individual's risk of disease. However as a group, overweight and obese individuals have an increased risk of many diseases.

They have a greatly increased risk of:

- Type 2 Diabetes
- Gall bladder disease.
- Hypertension.
- Dyslipidaemia
- Insulin Resistance
- Atherosclerosis
- Sleep apnoea
- Breathlessness
- Asthma
- Social isolation and depression
- Daytime sleepiness and fatigue.

They have a moderately increased risk of:

- Cardiovascular diseases(i.e. stroke, heart attack).
- Gout/hyperuricaemia.
- Osteoarthritis.
- Respiratory disease.
- Hernia.
- Psychological problems.

They have a slightly increased risk of:

- Some forms of cancer (breast, colon and endometrial cancers).
- Reproductive abnormalities.

- Impaired fertility.
- Polyscystic ovarian syndrome.
- Skin complications.
- Contaract.
- Varicose veins.
- Musculskeletal problems.
- Bad back.
- Stress incontinence.
- Oedema/cellulitis.

Underweight

Individuals who are underweight may be malnourished. In addition they have an increased risk of developing health problems including: - Compromised immune function with increased susceptibility to infections; - Anaemia; Osteoporosis - Menstrual irregularities - Impaired fertility.

2.2 Exercise

Exercise is an bodily activity that enhances or maintains physical fitness and overall health and wellness. It is performed for various reasons, including increasing growth and development, preventing aging strengthening muscles and the cardiovascular system, honingathletic skills, weight lossor maintenance, improving health and also for enjoyment. Many individuals choose to exercise outdoors where they can congregate in groups, socialize, and enhance well-being.

2.2.1 Classification

Physical exercises are generally grouped into three types, depending on the overall effect they have on the human body:

- Aerobic exercise is any physical activity that uses large muscle groups and causes the body to use more oxygen than it would while resting. The goal of aerobic exercise is to increase cardiovascular Endurance. Examples of aerobic exercise include running, cycling and swimming brisk walking, skipping rope, rowing, hiking, playing tennis, continuous training, and long slow distance training.
- Anaerobic exercise which includes strength and resistance training, can firm, strengthen, and tone muscles, as well as improve bone strength, balance and coordination.Examples of strength moves are push-ups, pull-ups,lunges, and biceps-curls using dumbbells.Anaerobic exercise also include weight training, functional training, eccentric training, interval training, sprinting and high-intensity interval training increase short-term muscle strength.
- Flexibility exercises stretch and lengthen muscles. Activities such as stretching help to improve joint flexibility and keep muscles limber.The goal is to improve the range of motion which can reduce the chance of injury. Physical exercise can also include training that focuses on accuracy, agility, power and speed. Sometimes the terms ‘dynamic’ and ‘static’ are used.’Dynamic’ exercises such as steady running, tend to produce a lowering of the Diastolic blood pressure during exercise, due to the improved blood flow. Conversely, static exercise (such as weight-lifting) can cause the systolic pressure to rise significantly, albeit transiently, during the performance of the exercise.

2.2.2 Health effects

Physical exercise is important for maintaining physical fitness and contribute to maintaining a healthy weight, regulating digestive health, building and maintaining healthy bone density, muscle strength, and joint mobility, promoting physiological well-being, reducing surgical risks, and strengthening the immune system.

Some studies indicate that exercise may increase life expectancy and the overall quality of life.People who participate in moderate to high levels of physical exercise have a lower mortality rate compared to individuals who by comparison are not physically active.Moderate levels of exercise have been correlated with preventing aging by reducing inflammatory potential. The majority of the benefits from exercise are achieved with around 3500 Metabolic equivalent (MET) minutes per week. For example, climbing stairs 10 minutes, vacuuming 15 minutes, gardening 20 minutes, running 20 minutes, and walking or bicycling for transportation 25 minutes on a daily basis would together achieve about 3000 MET minutes a week. A lack of physical activity causes approximately 6% of the burden of disease from coronary heart disease, 7% of type 2 diabetes, 10% of breast cancer and 10% of colon cancer worldwide. Overall, physical inactivity causes 9% of premature mortality worldwide [5].

2.3 Sleep

Sleep is a naturally recurring state of mind and body, characterized by altered consciousness relatively inhibited sensory activity, inhibition of nearly all voluntary muscles and reduced interactions with surroundings. It is distinguished from wakefulness by a decreased ability to react to stimuli, but more reactive than Coma or disorder of consciousness, sleep displaying very different and active brain patterns. Sleep occurs in repeated periods in which the body alternates between two distinct modes: REM sleep and non-REM sleep. Although REM stands for "rapid eye movement", this mode of sleep has many other aspects, including virtual Paralysis of the body. A well-known feature of sleep is the dream a, an experience typically recounted in narrative form, which resembles waking life while in progress, but which usually can later be distinguished as fantasy. During sleep, most of the body's system are in an anabolic state, helping to restore the immune, nervous, skeletal, and muscular systems; these are vital processes that maintain mood, memory, and cognitive function, and play a large role in the function of the endocrine and Immune systems The internal Circadian clock promotes sleep daily at night. The diverse purposes and mechanisms of sleep are the subject of substantial ongoing research. Sleep is a highly conserved behavior across animal evolution. Humans may suffer from various sleep disorders including Dyssomnia as such as insomnia hypersomnia, narcolepsy and Sleep apnea, Parasomnia such as Sleep walking and REM behaviour disorder, bruxism and Circadian rhythm sleep disorders. The advent of artificial light has substantially altered sleep timing in industrialized countries.

2.3.1 Physiology

The most pronounced physiological changes in sleep occur in the brain. The brain uses significantly less energy during sleep than it does when awake, especially during non-REM sleep. In areas with reduced activity, the brain restores its supply of adenosine triphosphate(ATP), the molecule used for short-term storage and transport of energy. In quiet waking, the brain is responsible for 20% of the body's energy use, thus this reduction has a noticeable effect on overall energy consumption. Sleep increases the sensory threshold. In other words, sleeping persons perceive fewer stimuli, but can generally still respond to loud noises and other salient sensory events. During slow-wave sleep, humans secrete bursts of growth hormone. All sleep, even during the day, is associated with secretion of prolactin. Key physiological methods for monitoring and measuring changes during sleep include electroencephalography (EEG) of brain waves, electrooculography (EOG) of eye movements, and electromyography (EMG) of skeletal muscle activity. Simultaneous collection of these measurements is called polysomnography, and can be performed in a specialized sleep laboratory. Sleep researchers also use simplified electrocardiography (EKG) for cardiac activity and actigraphy for motor movements.

2.3.2 Awakening

Awakening can mean the end of sleep, or simply a moment to survey the environment and readjust body position before falling back asleep. Sleepers typically awaken soon after the end of a REM phase or sometimes in the middle of REM. Internal circadian indicators, along with successful reduction of homeostatic sleep need, typically bring about awakening and the end of the sleep cycle. Awakening involves heightened electrical activation in the brain, beginning with the thalamus and spreading throughout the cortex. During a night's sleep, a small amount of time is usually spent in a waking state. As measured by electroencephalography, young females are awake for 0–1% of the larger sleeping period; young males are awake for 0–2%. In adults, wakefulness increases, especially in later cycles. One study found 3% awake time in the first ninety-minute sleep cycle, 8% in the second, 10% in the third, 12% in the fourth, and 13–14% in the fifth. Most of this awake time occurred shortly after REM sleep. Today, many humans wake up with an alarm clock; however, people can also reliably wake themselves up at a specific time with no need for an alarm.Many sleep quite differently on workdays versus days off, a pattern which can lead to chronic circadian desynchronization. Many people regularly look at television and other screens before going to bed, a factor which may exacerbate disruption of the circadian cycle. Scientific studies on sleep have shown that sleep stage at awakening is an important factor in amplifying sleep inertia.

2.3.3 Social timing

Humans are also influenced by aspects of social time, such as the hours when other people are awake, the hours when work is required, the time on the clock, etc. Time zones, standard times used to unify the timing for people in the same area, correspond only approximately to the natural rising and setting of the sun. The approximate nature of the timezone can be shown with China, a country which used to span five time zones and now officially uses only one (UTC +8).

2.3.4Distribution

In poly-phasic sleep, an organism sleeps several times in a 24-hour cycle. Mono-phasic sleep occurs all at once. Under experimental conditions, humans tend to alternate more frequently between sleep and wakefulness (i.e., exhibit more poly-phasic sleep) if they have nothing better to do. Given a 14-hour period of darkness in experimental conditions, humans tended towards bi-modal sleep, with two sleep periods concentrated at the beginning and at the end of the dark time. Bi-modal sleep in humans was more common before the industrial revolution. Different characteristic sleep patterns, such as the familiarly so-called "early bird" and "night owl", are called chronotypes. Genetics and sex have some influence on chronotype, but so do habits. Chronotype is also liable to change over the course of a person's lifetime. Seven-year-olds are better disposed to wake up early in the morning than are fifteen-year-olds. Chronotypes far outside the normal range are called circadian rhythm sleep disorders.

2.3.5. Genetics

It is hypothesized that a considerable amount of sleep-related behavior, such as when and how long a person needs to sleep, is regulated by genetics. Researchers have discovered some evidence that seems to support this assumption. Monozygotic (identical) but not dizygotic (fraternal) twins tend to have similar sleep habits. Neurotransmitters, molecules whose production can be traced to specific genes, are one genetic influence on sleep which can be analyzed. And the circadian clock has its own set of genes. Genes which may influence sleep include ABCC9, DEC2, and variants near PAX 8 and VRK2.

2.3.6. Recommended Sleep duration.

Children need many hours of sleep per day in order to develop and function properly: up to 18 hours for newborn babies, with a declining rate as a child ages. Early in 2015, after a two-year study, the National Sleep Foundation in the US announced newly revised recommendations as shown in the table below.

Recommended sleeping hours

Age and Condition	Sleep Needs
Newborns (0-3 months)	14 to 17 hours
Infants (4-11 months)	12 to 15 hours
Toddlers (1-2 years)	11 to 14 hours
Preschoolers (3-4 years)	10 to 13 hours
School-age (5-12 years)	9 to 11 hours
Teenagers (13-17 years)	8 to 10 hours
Adults (18-64 years)	7 to 9 hours
Older Adults (65 years and over)	7 to 8 hours

2.3.7Disorders

1.Insomnia

Insomnia is a general term for difficulty falling asleep and/or staying asleep. Insomnia is the most common sleep problem, with many adults reporting occasional insomnia, and 10–15% reporting a chronic condition. Insomnia can have many different causes, including psychological stress, a poor sleep environment, an inconsistent sleep schedule, or excessive mental or physical stimulation in the hours before bedtime. Insomnia is often treated through behavioral changes like keeping a regular sleep schedule, avoiding stimulating or stressful activities before bedtime, and cutting down on stimulants such as caffeine. The sleep environment may be improved by installing heavy drapes to shut out all sunlight, and keeping computers, televisions and work materials out of the sleeping area. A 2010 review of published scientific research suggested that exercise generally improves sleep for most people, and helps sleep disorders such as insomnia. The optimum time to exercise may be 4 to 8 hours before bedtime, though exercise at any time of day is beneficial, with the exception of heavy exercise taken shortly before bedtime, which may disturb sleep. However, there is insufficient evidence to draw detailed conclusions about the relationship between exercise and sleep. Sleeping medications such as Ambien and Lunesta are an increasingly popular treatment for insomnia. Although these nonbenzodiazepine medications are generally believed to be better and safer than earlier generations of sedatives, they have still generated some controversy and discussion regarding sideeffects. White noise appears to be a promising treatment for insomnia.

2.Obstructive sleep apnea

Obstructive sleep apnea is a condition in which major pauses in breathing occur during sleep, disrupting the normal progression of sleep and often causing other more severe health problems. Apneas occur when the muscles around the patient's airway relax during sleep, causing the airway to collapse and block the intake of oxygen. Obstructive sleep apnea is more common than central sleep apnea. As oxygen levels in the blood drop, the patient then comes out of deep sleep in order to resume breathing. When several of these episodes occur per hour, sleep apnea rises to a level of seriousness that may require treatment. Diagnosing sleep apnea usually requires a professional sleep study performed in a sleep clinic, because the episodes of wakefulness caused by the disorder are extremely brief and patients usually do not remember experiencing them. Instead, many patients simply feel tired after getting several hours of sleep and have no idea why. Major risk factors for sleep apnea include chronic fatigue, old age, obesity and snoring.

3.Other disorders

Sleep disorders include narcolepsy, periodic limb movement disorder (PLMD), restless leg syndrome (RLS), upper airway resistance syndrome (UARS), and the circadian rhythm sleep disorders. Fatal familial insomnia, or FFI, an extremely rare genetic disease with no known treatment or cure, is characterized by increasing insomnia as one of its symptoms; ultimately sufferers of the disease stop sleeping entirely, before dying of the disease. Somnambulism, known as sleep walking, is also a common sleeping disorder, especially among children. In somnambulism the individual gets up from his/her sleep and wanders around while still sleeping. Older people may be more easily awakened by disturbances in the environment may to some degree lose the ability to consolidate sleep.

2.4.1 Environment

Humans interact with the environment constantly. These interactions affect quality of life, years of healthy life lived, and health disparities. The World Health Organization (WHO) defines environment, as it relates to health, as "all the physical, chemical, and biological factors external to a person, and all the related behaviors." Environmental health consists of preventing or controlling disease, injury, and disability related to the interactions between people and their environment. The Healthy People 2020 Environmental Health objectives focus on 6 themes, each of which highlights an element of environmental health: 1. Outdoor air quality 2. Surface and ground water quality 3. Toxic substances and hazardous wastes 4. Homes and communities 5. Infrastructure and surveillance 6. Global environmental health Creating healthy environments can be complex and relies on continuing research to better understand the effects of exposure to environmental hazards on people's health.

2.4.1 Why Is Environmental Health Important?

Maintaining a healthy environment is central to increasing quality of life and years of healthy life. Globally, 23% of all deaths and 26% of deaths among children under age 5 are due to preventable environmental factors.[1] Environmental factors are diverse and far reaching. They include: - Exposure to hazardous substances in the air, water, soil, and food - Natural and technological disasters - Climate change - Occupational hazards - The built environment

Understanding Environmental Health The 6 themes of the Environmental Health topic area draw attention to elements of the environment and their linkages to health.

Outdoor Air Quality

Poor air quality is linked to premature death, cancer, and long-term damage to respiratory and cardiovascular systems. Progress has been made to reduce unhealthy air emissions, but in 2008, approximately 127 million people lived in U.S. counties that exceeded national air quality standards. Decreasing air pollution is an important step in creating a healthy environment.

Surface and Ground Water

Surface and ground water quality concerns apply to both drinking water and recreational waters. Contamination by infectious agents or chemicals can cause mild to severe illness. Protecting water sources and minimizing exposure to contaminated water sources are important parts of environmental health.

Toxic Substances and Hazardous Wastes

The health effects of toxic substances and hazardous wastes are not yet fully understood. Research to better understand how these exposures may impact health is ongoing. Meanwhile, efforts to reduce exposures continue. Reducing exposure to toxic substances and hazardous wastes is fundamental to environmental health.

Homes and Communities

People spend most of their time at home, work, or school. Some of these environments may expose people to: - Indoor air pollution - Inadequate heating and sanitation - Structural problems - Electrical and fire hazards - Lead-based paint hazards

Infrastructure and Surveillance

Preventing exposure to environmental hazards relies on many partners, including state and local health departments. Personnel, surveillance systems, and education are important resources for investigating and responding to disease, monitoring for hazards, and educating the public. Additional methods and greater capacity to measure and respond to environmental hazards are needed.

Global Environmental Health

Water quality is an important global challenge. Diseases can be reduced by improving water quality and sanitation and increasing access to adequate water and sanitation facilities.

Emerging Issues in Environmental Health

Environmental health is a dynamic and evolving field. While not all complex environmental issues can be predicted, some known emerging issues in the field include:

Climate Change

Climate change is projected to impact sea level, patterns of infectious disease, air quality, and the severity of natural disasters such as floods, droughts, and storms.

Disaster Preparedness

Preparedness for the environmental impact of natural disasters as well as disasters of human origin includes planning for human health needs and the impact on public infrastructure, such as water and roadways.

Nanotechnology

The potential impact of nanotechnology is significant and offers possible improvements to: - Disease prevention, detection, and treatment - Electronics - Clean energy - Manufacturing - Environmental risk assessment

Objective of this study:

1. Our main objective is to elicit the information that what are the major factors that affect students' health.
2. How can we keep ourselves healthy by considering some points and getting knowledge about the health issues.
3. How many students lie in different category of BMI like (underweight, healthy, overweight, obese, highly obese). 4.Parameters that we have selected for the study, do they have some significant effect on the health combined.

CHAPTER 2

METHODOLOGY

2.1 POPULATION

The word population is used to denote the aggregate from which the sample is chosen. Population means all the members that meet a set of specification or a specified criterion.For example, the population of the Mohd Habib Hall is defined as all the students residing in the hall.

2.2 SAMPLE

A finite subset of statistical individuals in a population is called a sample and the number of individuals in a sample is called the sample size.

2.2.1 Types of Sampling:

Some of the commonly known and frequently used types of sampling are :

a. Purposive sampling
b. Random sampling
c. Stratified sampling
d. Systematic Sampling

2.3 STRATIFIED SAMPLING

Where the entire heterogeneous population is divided into a number of homogeneous groups, usually termed as strata, which differ from one another but each of these groups is homogenous within itself. Then units are sampled at random from each of these stratum, the sample size in each stratum varies according to the relative importance of the stratum in the population. The sample, which is the aggregate of the sampled units of each of the stratum, is termed as stratified sampling and the technique of drawing this sample is known as stratified sampling. Such a sample is by far the best and can safely be considered as representative of the population from which it has been drawn. In our project , we used stratified sampling. In the total population of M.H.HALL , we have selected sample randomly in the three strata named as U.G., P.G., Ph.D.

2.4 SAMPLE SURVEY

A sample survey is a process for collecting data on a sample of observations which are selected from the population of interest using a probability-based sample design. In sample surveys, certain methods are often used to improve the precision and control the costs of survey data collection.

2.5 DATA TYPES

Having a good understanding of the different data types, also called measurement scales, is a crucial prerequisite for doing Exploratory Data Analysis (EDA), since one can use certain statistical measurements only for specific data types. One also needs to know which data type you are dealing with to choose the right visualization method. Think of data types as a way to categorize different types of variables. We will discuss the main types of variables and look at an example for each. We will sometimes refer to them as measurement scales.

2.5.1 CATEGORICAL DATA

Categorical data represents characteristics. Therefore it can represent things like a person's gender, language etc. Categorical data can also take on numerical values (Example: 1 for female and 0 for male). Note that those numbers don't have mathematical meaning.

2.5.2 NOMINAL DATA

Nominal values represent discrete units and are used to label variables, that have no quantitative value. Just think of them as labels. Note that nominal data that has no order. Therefore if you would change the order of its values, the meaning would not change.

2.5.3 ORDINAL DATA

Ordinal values represent discrete and ordered units. It is therefore nearly the same as nominal data, except that it's ordering matters. Note that the difference between Elementary and High School is different than the difference between High School and College. This is the main limitation of ordinal data, the differences between the values is not really known. Because of that, ordinal scales are usually used to measure non-numeric features like happiness, customer satisfaction and so on. #### **2.5.4 NUMERICAL DATA**

2.5.4.1 DISCRETE DATA

We speak of discrete data if its values are distinct and separate. In other words: We speak of discrete data if the data can only take on certain values. This type of data can't be measured but it can be counted. It basically represents information that can be categorized into a classification. An example is the number of heads in 100 coin flips. We can check by asking the following two questions whether you are dealing with discrete data or not: Can you count it and can it be divided up into smaller and smaller parts? On the contrary, if the data could be measured but not counted, we would speak of continuous data.

2.5.4.2 CONTINUOUS DATA

Continuous Data represents measurements and therefore their values can't be counted but they can be measured. An example would be the height of a person, which you can describe by using intervals on the real number line.

2.5.4.3 INTERVAL DATA

Interval values represent ordered units that have the same difference. Therefore we speak of interval data when we have a variable that contains numeric values that are ordered and where we know the exact differences between the values. The problem with interval values data is that they don't have a „true zero“. That means in regards to our example, that there is no such thing as no temperature. With interval data, we can add and subtract, but we cannot multiply, divide or calculate ratios. Because there is no true zero, a lot of descriptive and inferential statistics can't be applied.

2.5.4.4 RATIO DATA

Ratio values are also ordered units that have the same difference. Ratio values are the same as interval values, with the difference that they do have an absolute zero. Good examples are height, weight, length etc. Why Data Types are important? Data types are an important concept because statistical methods can only be used with certain data types. You have to analyze continuous data differently than categorical data otherwise it would result in a wrong analysis. Therefore knowing the types of data you are dealing with, enables you to choose the correct method of analysis.

2.6 Statistical Methods

2.6.1 Nominal Data

When you are dealing with nominal data, you collect information through: Frequencies: The Frequency is the rate at which something occurs over a period of time or within a dataset. Proportion: You can easily calculate the proportion by dividing the frequency by the total number of events. (e.g how often something happened divided by how often it could happen) Percentage: This one doesn't need an explanation. Visualization Methods: To visualize nominal data you can use a pie chart or a bar chart.

2.6.2 Ordinal Data

When you are dealing with ordinal data, you can use the same methods like with nominal data, but you also have access to some additional tools. Therefore you can summarize your ordinal data with frequencies, proportions, percentages. And you can visualize it with pie and bar charts. Additionally, you can use percentiles, median, mode and the inter-quartile range to summarize your data.

2.6.3 Continuous Data

When you are dealing with continuous data, you can use the most methods to describe your data. You can summarize your data using percentiles, median, inter-quartile range, mean, mode, standard deviation, and range. Visualization Methods: To visualize continuous data, we can use a histogram or a box-plot. With a histogram, we can check the central tendency, variability and kurtosis of a distribution. Note that a histogram can't show an outlier. This is why we also use box-plots.

2.7 SUMMARY

Data collection

Data collection is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes. Data collection is a component of research in all fields of study including physicaland social sciences, humanities, and business. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same. The goal for all data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the questions that have been posed.

Importance:

Regardless of the field of study or preference for defining data (quantitative or qualitative), accurate data collection is essential to maintaining the integrity of research. The selection of appropriate data collection instruments (existing, modified, or newly developed) and clearly delineated instructions for their correct use reduce the likelihood of errors. A formal data collection process is necessary as it ensures that the data gathered are both defined and accurate. This way, subsequent decisions based on arguments embodied in the findings are made using valid data. The process provides both a baseline from which to measure and in certain cases an indication of what to improve.

Impact of faulty data:

Consequences from improperly collected data include: - Inability to answer research questions accurately; - Inability to repeat and validate the study.Distorted findings result in wasted resources and can mislead other researchers into pursuing fruitless avenues of investigation; it may also compromise decisions, for example for public policy, which may cause disproportionate harm.

2.8 Different types of data:

Data can be classified as either primary or secondary.

2.8.1 Primary Data:

Primary data means original data that has been collected specially for the purpose in mind. It means someone collected the data from the original source first hand. Data collected this way is called primary data.

The people who gather primary data may be an authorized organization, investigator, enumerator or they may be just someone with a clipboard. Those who gather primary data may have knowledge of the study and may be motivated to make the study a success. These people are acting as a witness so primary data is only considered as reliable as the people who gathered it. Research where one gathers this kind of data is referred to as field research.

Merits:

1. Degree of accuracy is quite high.
2. It does not require extra caution.
3. It depicts the data in great detail.
4. Primary source of data collection frequently includes definitions of various! and units used.
5. For some investigations, secondary data are not available.

Demerits:

1. Collection of data requires a lot of time.
2. It requires lot of finance.
3. In some enquiries it is not possible to collect primary data.
4. It requires a lot of labor.
5. It requires a lot of skill.

Methods of data collection of Primary data:

Direct personal investigation: This is a method in which the investigation is done personally for the required data. Interview/questionnaires: Under this method the investigator collects the data from the respondents putting questions to them regarding required data. Discussion with community leaders: Some data which are required cannot be collected through personal investigation or through interview so community leaders are approached to fetch information for the required data.

2.9 Hypothesis Testing:

Hypothesis testing was introduced by Ronald Fisher, Jerzy Neyman, Karl Pearson and Pearson's son, Egon Pearson. Hypothesis testing is a statistical method that is used in making statistical decisions using experimental data. Hypothesis Testing is basically an assumption that we make about the population parameter.

Key terms and concepts:

Null Hypothesis:

Null hypothesis is a statistical hypothesis that assumes that the observation is due to a chance factor. Null hypothesis is denoted by $H_0: \mu_1 = \mu_2$, which shows that there is no difference between the two population means.

Alternative Hypothesis:

Contrary to the null hypothesis, the alternative hypothesis shows that observations are the result of a real effect.

Level of significance: Refers to the degree of significance in which we accept or reject the null-hypothesis. 100% accuracy is not possible for accepting or rejecting a hypothesis, so we therefore select a level of significance that is usually 5%.

Type I error:

When we reject the null hypothesis, although that hypothesis was true Type I error is denoted by . In hypothesis testing, the normal curve that shows the critical region is called the region.

Type II errors:

When we accept the null hypothesis but it is false. Type II errors are denoted by . In Hypothesis testing, the normal curve that shows the acceptance region is called the region.

Power:

Usually known as the probability of correctly accepting the null hypothesis, $1-\beta$ is called power of the analysis.

One-tailed test:

When the given statistical hypothesis is one value like $H_0 : \mu_1 = \mu_2$, it is called the one-tailed test.

Two-tailed test:

When the given statistics hypothesis assumes a less than or greater than value, it is called the two-tailed test.

Statistical decision for hypothesis testing:

In statistical analysis, we have to make decisions about the hypothesis. These decisions include deciding if we should accept the null hypothesis or if we should reject the null hypothesis. Every test in hypothesis testing produces the significance value for that particular test. In Hypothesis testing, if the significance value of the test is greater than the predetermined significance level, then we accept the null hypothesis. If the significance value is less than the predetermined value, then we should reject the null hypothesis. For example, if we want to see the degree of relationship between two stock prices and the significance value of the correlation coefficient is greater than the predetermined significance level, then we can accept the null hypothesis and conclude that there was no relationship between the two stock prices. However, due to the chance factor, it shows a relationship between the variables.

2.9.1 Z-test

This procedure calculates sample size and statistical power for testing a single proportion using either the exact test or other approximate z-tests. … When the sample sizes are small or the proportions are extreme (i.e. less than 0.2 or greater than 0.8) the binomial calculations are much more accurate. In a z-test, the sample is assumed to be normally distributed. A z-score is calculated with population parameters such as "population mean" and "population standard deviation" and is used to validate a hypothesis that the sample drawn belongs to the same population.

Null Hypothesis: Sample mean is same as the population mean

Alternate Hypothesis: Sample mean is not same as the population mean

The statistics used for this hypothesis testing is called z-statistic, the score for which is calculated as $z = \frac{(x—\mu)}{(\sigma/\sqrt{n})}$, where

x = sample mean

= population mean

$\frac{\sigma}{\sqrt{n}}$ = population standard deviation

If the test statistic is lower than the critical value, accept the hypothesis or else reject the hypothesis

Test for Single Proportion.

The test statistic is a z-score (z) defined by the following equation. $z=\frac{(p-P)}{\sigma}$ where P is the hypothesized value of population proportion in the null hypothesis, p is the sample proportion, and is the standard deviation of the sampling distribution.

Test Statistics is defined and given by the following function: Point estimate We estimate the proportion, p, as:

where x is the number in the sample who have the trait or outcome of interest, and n is the size of the sample. Hypothesis Tests:

- Null hypothesis $H_0 : p = p_0$
- Alternative Hypothesis $H_1: p \neq p_0$

This hypothesis considers whether the population proportion is equivalent to some pre-specified value, p_0. This value might be of historical interest or a result obtained in another study that we are trying to corroborate with our study data. A rule of thumb used to perform this test is that both np_0 and $n(1-p_0)$ are greater than five.

To perform this test, we: 1. Estimate the population proportion by the sample proportion. 2. Calculate the following test statistic, which under the null hypothesis, follows approximately (dependent on the rule of thumb stated above) a Standard

Normal Distribution: Formula:-
$z = \frac{p-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Where,

z= Test statistics

n= Sample size

p_o = Null hypothesized value

p = Observed proportion

where n is the sample size.

Decision Rule:

Reject if $Z > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the $\frac{1-\alpha}{2}$ percentile of the standard normal distribution

2.9.2Chi-Square Test

Chi-square test is used to compare categorical variables. There are two type of chi-square test 1. Goodness of fit test, which determines if a sample matches the population. 2. A chi-square fit test for two independent variables is used to compare two variables in a contingency table to check if the data fits. + A small chi-square value means that data fits + A high chi-square value means that data doesn't fit.

The hypothesis being tested for chi-square is

Null: Variable A and Variable B are independent

Alternate: Variable A and Variable B are not independent.

The statistic used to measure significance, in this case, is called chi-square statistic. The formula used for calculating the statistic is

$$\tilde{\chi}^2=\frac{1}{d}\sum\frac{(O_i - E_i)^2}{E_i})$$

where $O_i.c$ = observed frequency count at level i of Variable A and level c of Variable B

$E_i.c$ = expected frequency count at level i of Variable A and level c of Variable B

2.10 ggplot2

ggplot2 is an R package for producing statistical, or data, graphics, but it is unlike most other graphics packages because it has a deep underlying grammar. This grammar, based on the Grammar of Graphics (Wilkinson, 2005), is composed of a set of independent components that can be composed in many different ways. This makes ggplot2 very powerful, because you are not limited to a set of pre-specified graphics, but you can create new graphics that are precisely tailored for your problem. This may sound overwhelming, but because there is a simple set of core principles and very few special cases, ggplot2 is also easy to learn (although it may take a little time to forget your preconceptions from other graphics tools). Practically, ggplot2 provides beautiful, hassle-free plots, that take care of fiddly details like drawing legends. The plots can be built up iteratively and edited later. A carefully chosen set of defaults means that most of the time you can produce a publication-quality graphic in seconds, but if you do have special formatting requirements, a comprehensive theming system makes it easy to do what you want. Instead of spending time making your graph look pretty, you can focus on creating a graph that best reveals the messages in your data.

What is the grammar of graphics?

Wilkinson (2005) created the grammar of graphics to describe the deep features that underlie all statistical graphics. The grammar of graphics is an answer to a question: what is a statistical graphic? The layered grammar of graphics (Wickham, 2009) builds on Wilkinson's grammar, focussing on the primacy of layers and adapting it for embedding within R. In brief, the grammar tells us that a statistical graphic is a mapping from data to aesthetic attributes (colour, shape, size) of geometric objects (points, lines, bars). The plot may also contain statistical transformations of the data and is drawn on a specific coordinate system. Faceting can be used to generate the same plot for different subsets of the dataset. It is the combination of these independent components that make up a graphic. As the book progresses, the formal grammar will be explained in increasing detail. The first description of the components follows below. It introduces some of the terminology that will be used throughout the book and outlines the basic responsibilities of each component. Don't worry if it doesn't all make sense right away: you will have many more opportunities to learn about all of the pieces and how they fit together. - The data that you want to visualise and a set of aesthetic mappings describing how variables in the data are mapped to aesthetic attributes that you can perceive. - Geometric objects, geoms for short, represent what you actually see on the plot: points, lines, polygons, etc. - Statistical transformations, stats for short, summarise data in many useful ways. For example, binning and counting observations to create a histogram, or summarising a 2d relationship with a linear model. Stats are optional, but very useful. - The scales map values in the data space to values in an aesthetic space, whether it be colour, or size, or shape. Scales draw a legend or axes, which provide an inverse mapping to make it possible to read the original data values from the graph. - A coordinate system, coord for short, describes how data coordinates are mapped to the plane of the graphic. It also provides axes and gridlines to make it possible to read the graph. We normally use a Cartesian coordinate system, but a number of others are available, including polar coordinates and map projections. - A faceting specification describes how to break up the data into subsets and how to display those subsets as small multiples. This is also known as conditioning or latticing/trellising.

2.10.1 General Syntax

For histogram

\(>ggplot(data="**",aes(x=var-name,fill=var-name))+labels(y="ylabel",x="xlabel")+geom-histogram(col="**",position="fill")+facet-grid(~"strata-name")+***\)

For bar-plot

\(>ggplot(data="**",aes(x=var-name,fill=var-name))+labels(y="ylabel",x="xlabel")+geom-bar(col="**",position="fill")+facet-grid(~"strata-name")+***\)

For scatterplot

\(>ggplot(data="**",aes(x=var-name,y=var-name))+labels(y="ylabel",x="xlabel")+geom-point(col="**",position="fill")+facet-grid(~"strata-name")+***\)

CHAPTER 3

COLLECTION AND ANALYSIS OF DATA

3.1 Title of the Survey

"Analyzing BMI and other parameters of health, resident at MOHAMMAD HABIB Hall,Aligarh Muslim University (AMU), Aligarh (India), about HABITS RELATED TO HEALTH".

Objectives:

a. Conduct sample survey to collect information about Health Habits of UG, PG, and research scholars residing at Mohammad Habib Hall AMU, Aligarh (India).
b. Analyze the data that has been collected in the survey by using charts like (Bar chart, Pie chart and ggplot2).
c. For the insights and getting better result use t-test, z-test & chi-squared test.

d. Present the conclusions which are obtained in the analysis of the survey.

3.2 Population under Study

In sample survey the population Mohammad Habib Hall consists of the Undergraduates, Postgraduates and Research scholars which is situated in Aligarh Muslim University, Aligarh (India). Mohammad Habib hall was established to commemorate Prof. Mohammad Habib in the year 1972. It has 375 rooms and three hostels which are listed below. - Chakravarti hostel - Umaruddin hostel - Haider khan hostel There are 644 students in these three hostels in total, if we divide this figure into hostel wise there are 240, 217 and 187 students in Chakravarti, Umaruddin and Haider khan respectively.

3.3 Sampling Frame Construction

It was easy to get the list of the students' resident in the Hall from the Provost's Office of M.H. Hall. - Such list of the students of all the hostels of M.H. Hall under study was used as sampling frame for the sample selection. - The list was updated by visiting the individual hostels. - Finally corrected list of all the students under study was used as workable sampling frame for the sample selection.

3.4 Sampling Technique

Stratified sampling was used with the following strata: - Undergraduate Students (such as B.Sc/BCA) - Post-graduate & Professional course students (such as M.Sc/ MCA/MBA) - Research scholars (students of Ph.D)

3.5 Sample Size:

The standard sample size determination procedure (see Chapter 2) could not be adopted as the constraint of time did not permit to conduct a pilot survey. Therefore, an ad hoc sample size, i.e. 19.78% ≈ 20% of the total population, was used. Total population size, =644 20% of ≈ 128

3.5.1 Strata wise sample size:

Table-1

Strata Wise sample size

Strata	**Total Students**	**Samples**
UG	294	59
PG	207	41
Research	141	28

For this we take the sample units from the hostels of M.H. Hall of AMU, Aligarh, to estimate the population proportion.

The sample units were obtained by using the R command sample() **For UG**

```
set.seed(1)
y=sample(294,59)
sort(y)
```

```
## [1]  13  14  20  22  23  25  29  33  37  39  40  42  44  45  48  70  75  79  84
## [20]  85  89 103 104 105 108 110 111 121 129 130 141 150 156 160 165 167 172 176
```

```
## [39] 187 193 198 206 209 213 217 221 228 230 248 255 257 260 263 265 270 274 277
## [58] 281 283
```

For PG

```
set.seed(1)
x=sample(207,41)
sort(x)
```

```
## [1]   7  14  20  21  33  34  37  42  43  44  51  68  70  73  74  79  84  85  87
## [20]  89 105 106 110 111 121 126 129 148 156 162 163 165 166 167 170 172 179 182
## [39] 187 195 197
```

For Ph.D

```
set.seed(1)
z=sample(141,28)
sort(z)
```

```
## [1]   7  14  21  33  34  35  37  43  44  51  68  70  73  74  79  84  85  89 101
## [20] 105 106 110 127 129 130 133 134 140
```

3.6 Questionnaire

- It consists of 27 questions seeking the views of respondents on the topic
- Four demographic questions are there in the beginning of the questionnaire, asking about the

1. Age
2. Father's occupation &
3. Place of birth
4. Religion of the respondents

3.7 Field work

The field work was carried out as detailed in Chapter 2. Since, it is a known fact that no two individuals are alike it became crystal clear during the survey that each respondent has the unique way of expressing his own views. Overall, the respondents were helpful and cooperative. They tried their best to make our task easier. Sometimes, due to non-availability of respondents and some other reasons we have had to visit the selected units again and again to get the required information from all the selected individuals.

3.8 Master Chart:

Master chart was replaced by a MS-Excel spreadsheet which is a rectangular array with 120 rows and 28 columns and each column represents a variable in the data object.Codes used in the Master Chart: the Master Chart:
Questionnaire details

demographic_ques	**Q20_Q22**	**Q_10_12_13_23_25_26**	**Q_5_11_17**
Hindu	Strongly Disagree	Yes	Never
Islam	Disagree	No	Occassionally
Sikh	Can’t say	•	Regularly
Christian	Agree	•	•
Others	Strongly Agree	•	•

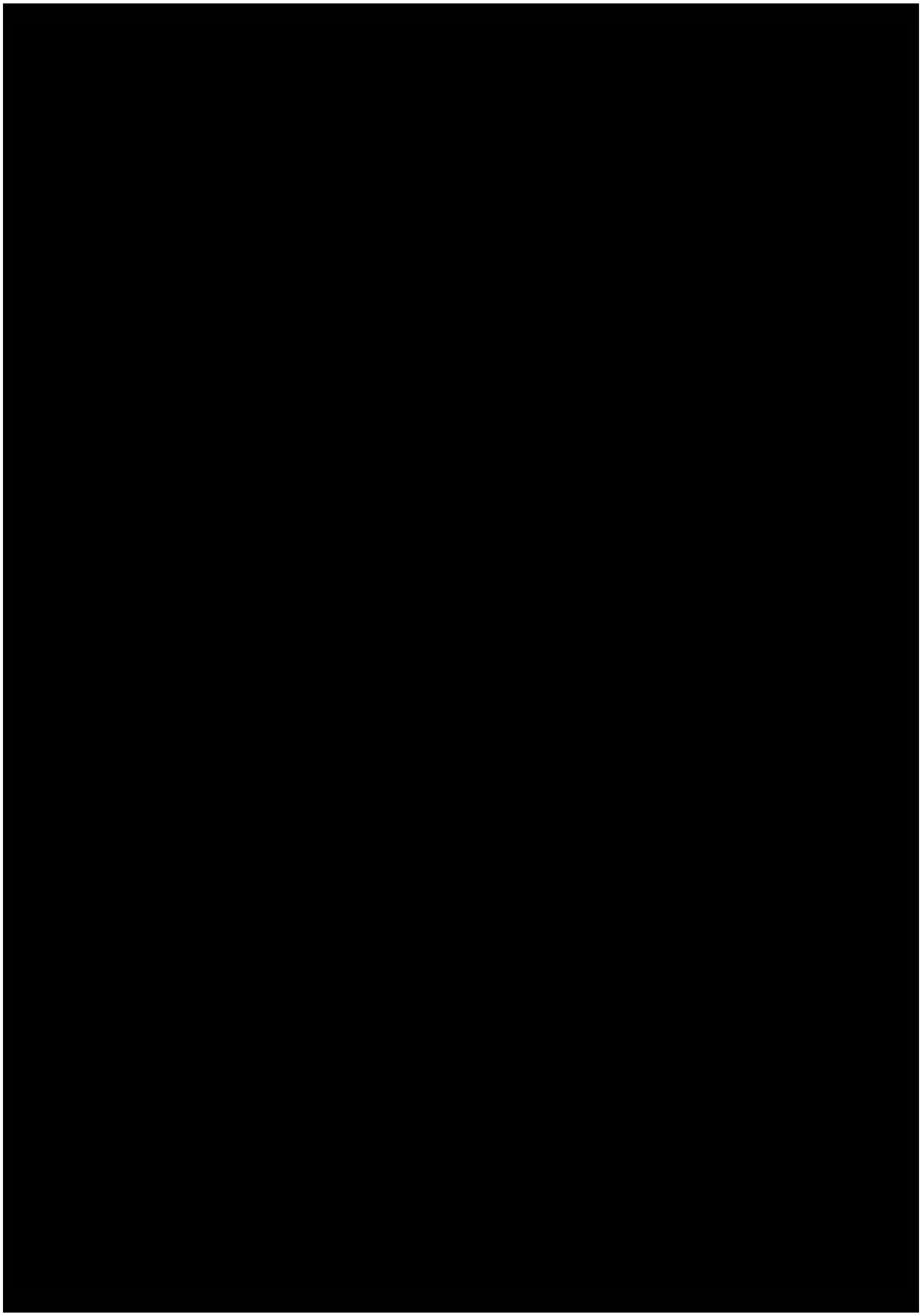
In question no. 1,2,3,4,6,7,8,9,13,14,15,16,18,19,24,27

Options 1,2,3,4… are used in respective manner

NA →No Response

Master Chart is given at the Appendix B

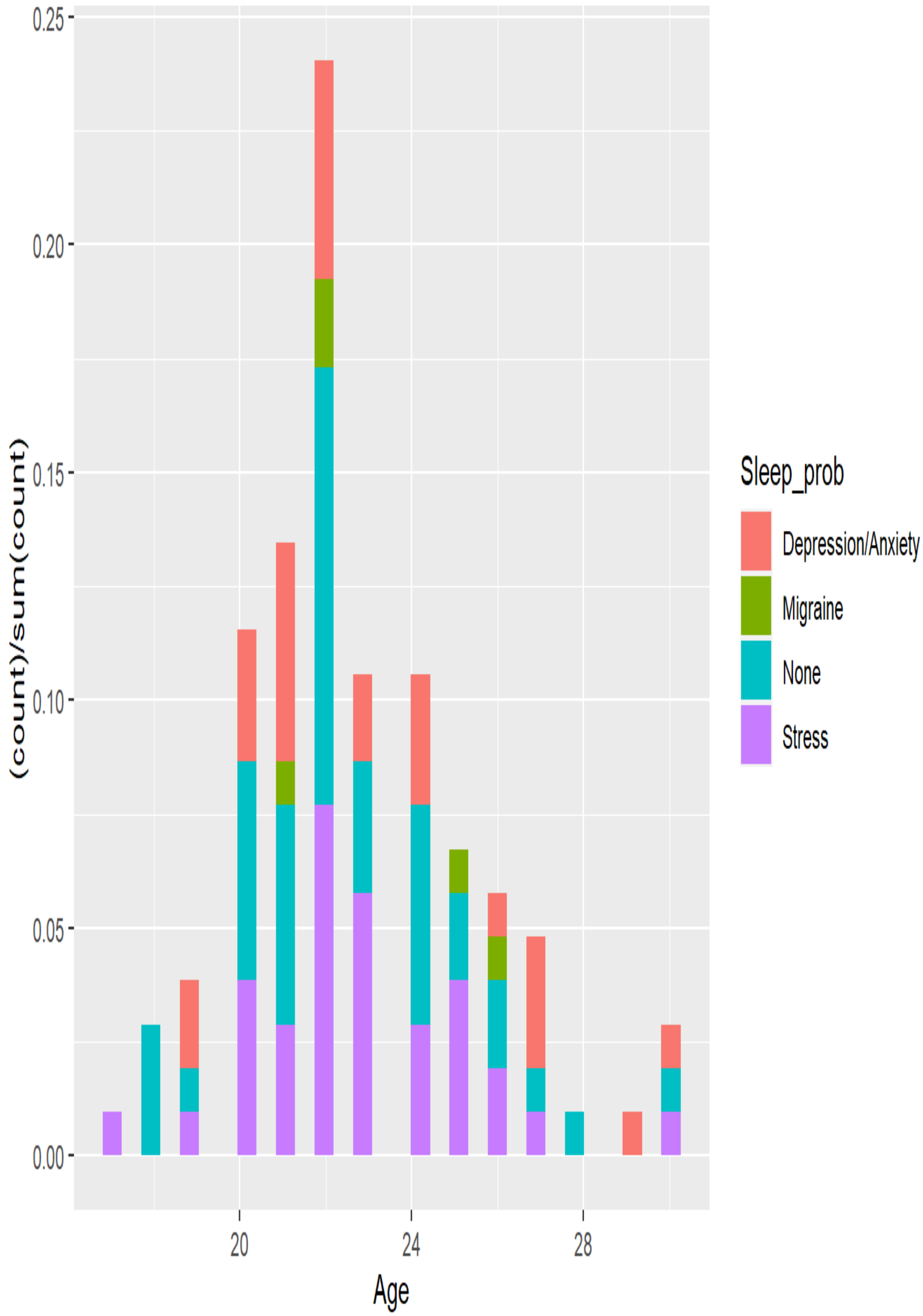
```
library("ggplot2")
#"ggplot2"
students=read.csv("Students.csv", header = T)
students<-na.omit(students)
ggplot(students)
```



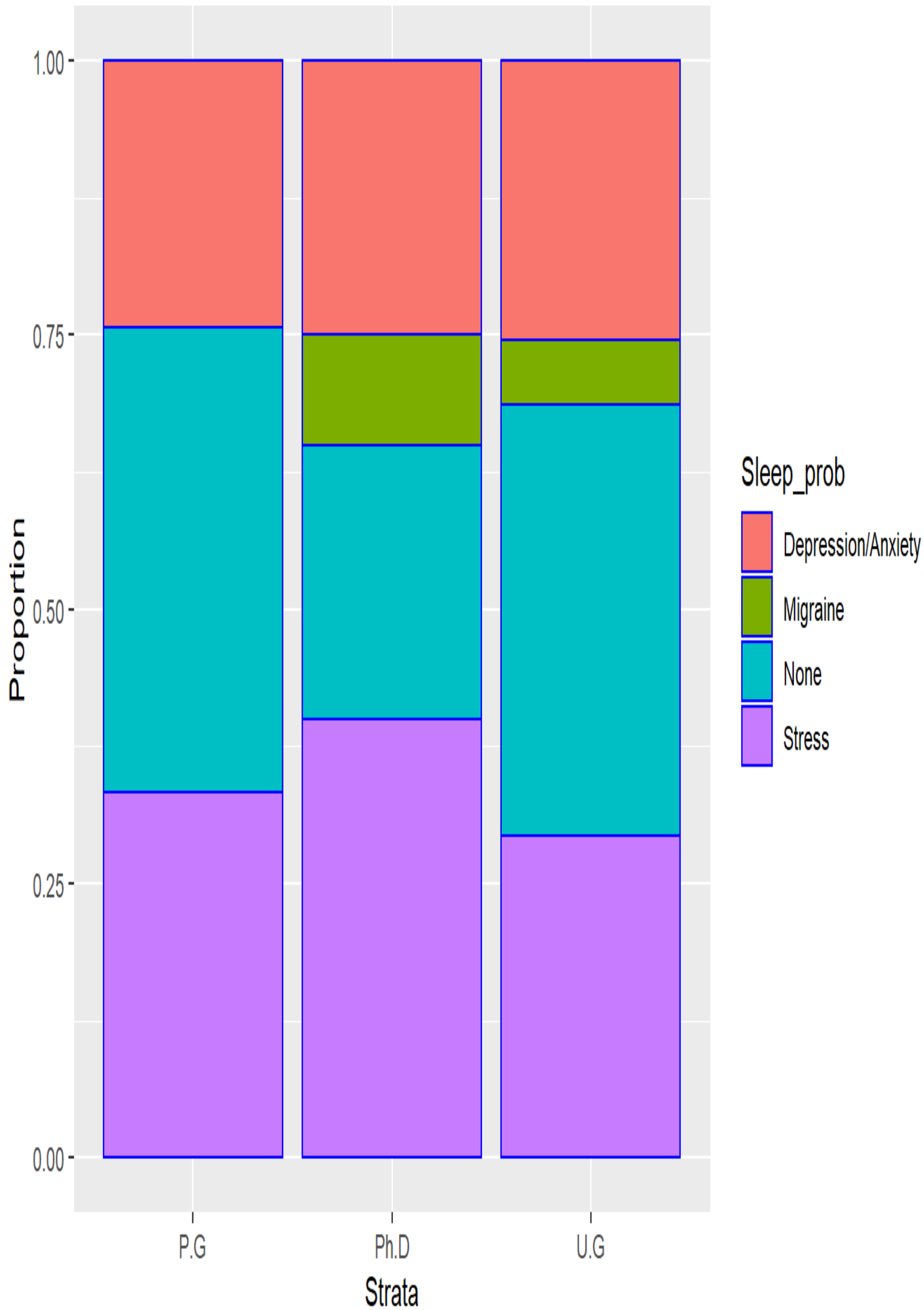
require(ggplot2)

#Analysis on Sleep_problems

ggplot(data=students,aes(x=Age,y=(..count..)/sum(..count..),fill=Sleep_prob))+geom_histogram(bins=30)




```
ggplot(data=students,aes(x=Strata,y=(..count..)/sum(..count..),fill=Sleep_prob))+geom_bar(position =
"fill",col="blue")+labs(y="Proportion")
```

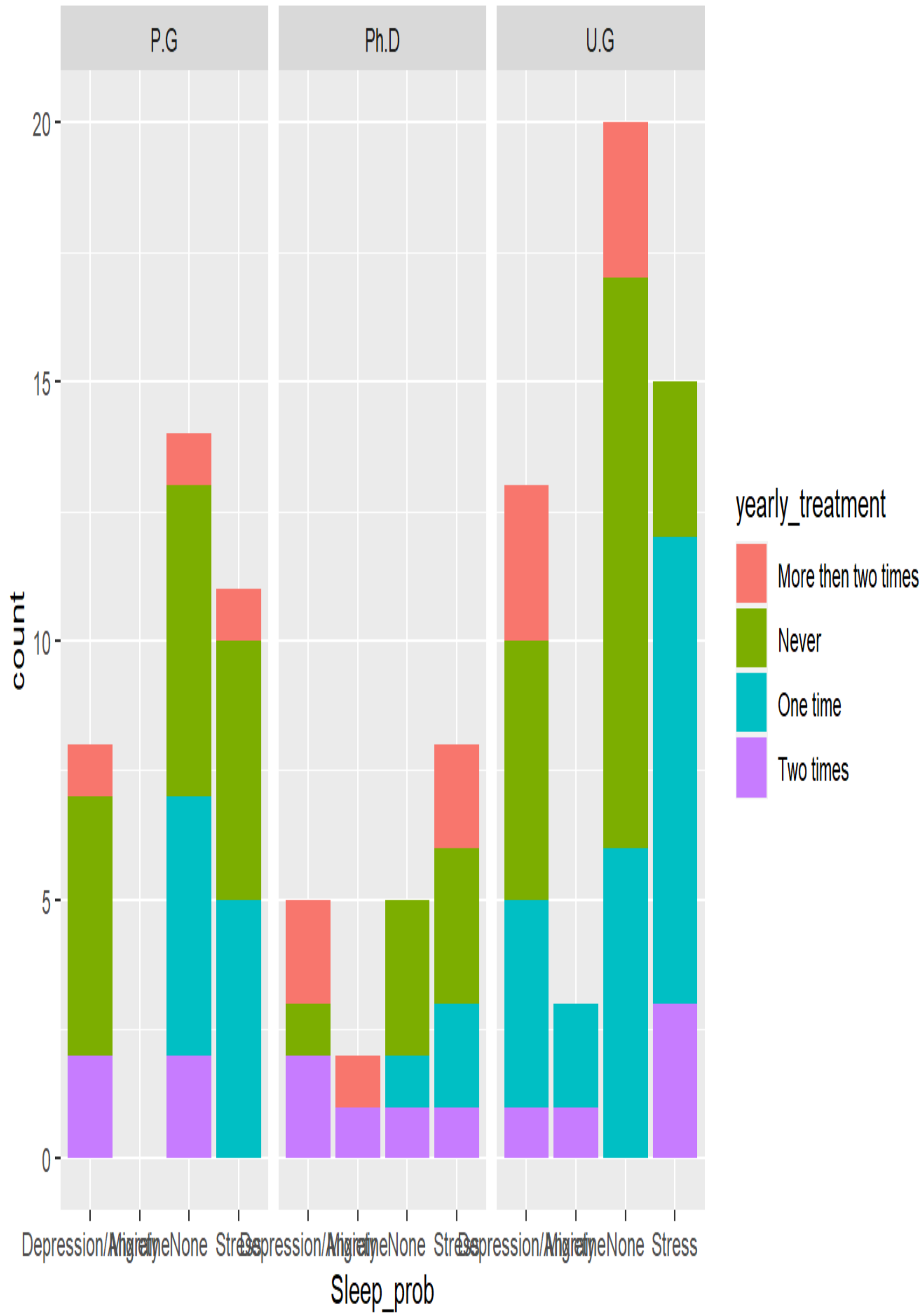


names(students)

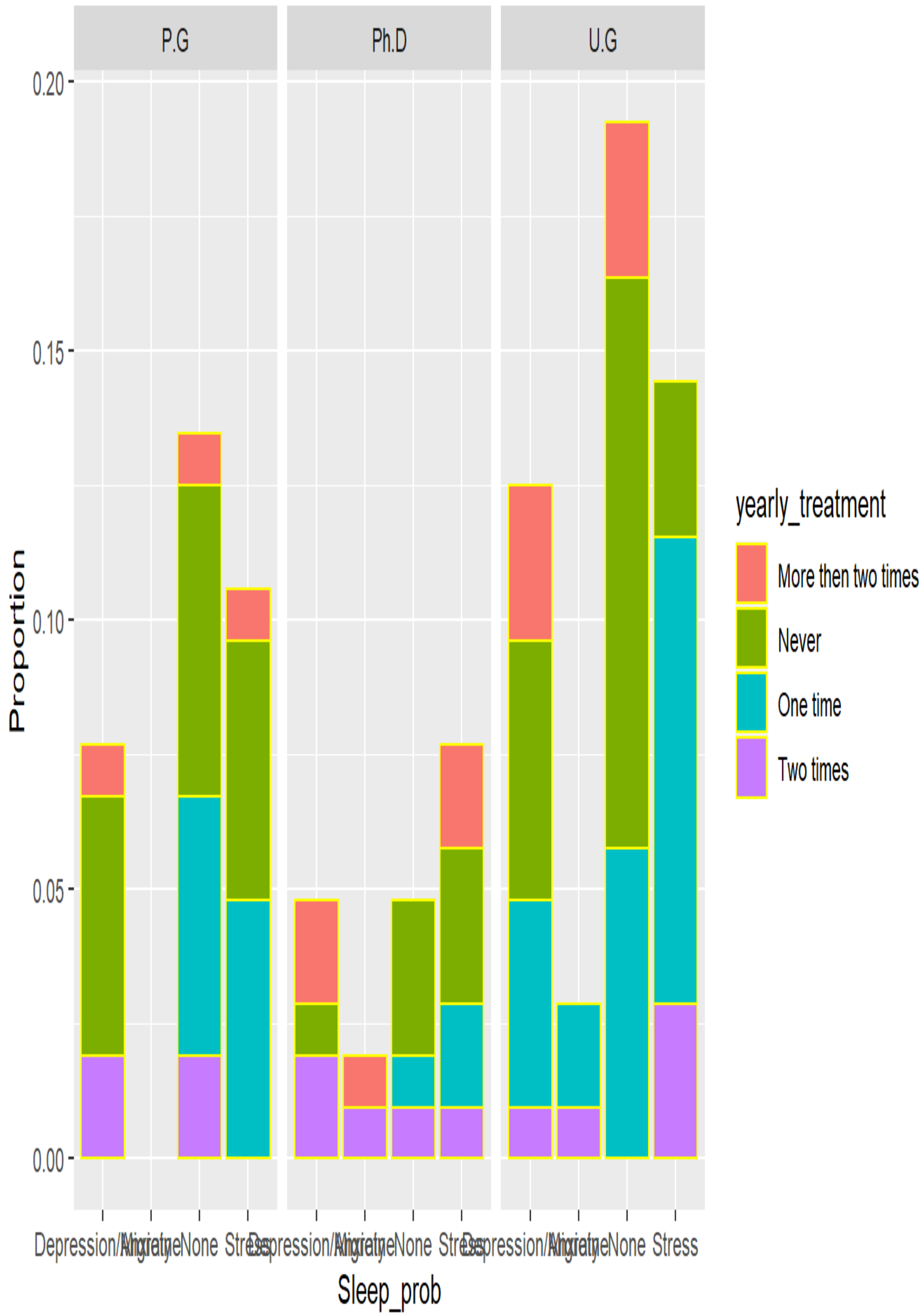
```
##  [1] "Strata"          "S.No."           "BMI"
##  [4] "Age"             "F.O."            "B.P."
##  [7] "Religion"        "Meal"            "Extra_diet"
## [10] "Aowater"         "Watquality"      "Smoke"
## [13] "Height"          "Weight"          "sleep_hrs"
## [16] "sleep_timing"    "Genetic_Sleep"   "Consume_T.C"
```

```
## [19] "Snoring"          "sleepinc"          "studyhrs"
## [22] "btofstudy"        "gadgetshrs"        "Headache"
## [25] "Sleep_prob"       "yearly_treatment"  "hhcndtn"
## [28] "x21"              "x22"               "Dust"
## [31] "x24"              "Participate_sport" "Exercise"
## [34] "Gym"
```

```
#Analysis on yearly treatment
ggplot(data=students,aes(x=Sleep_prob,fill=yearly_treatment))+geom_bar()+facet_grid(~Strata)
```

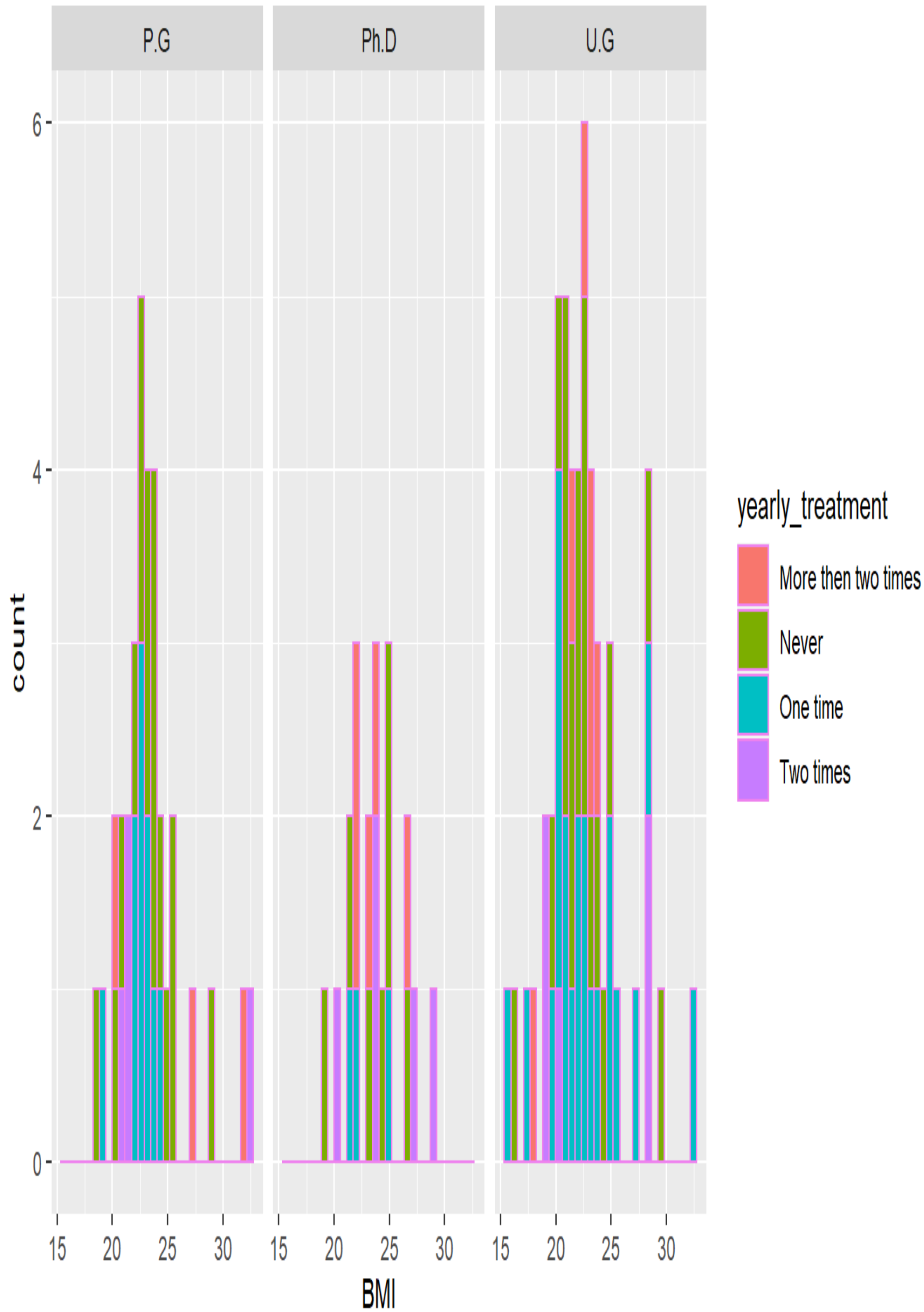



```
ggplot(data=students,aes(x=Sleep_prob,y=(..count..)/sum(..count..),fill=yearly_treatment))
+labs(y="Proportion")+geom_bar(col="yellow")+facet_grid(~Strata)
```



ggplot(data=students,aes(x=BMI,fill=yearly_treatment))+geom_histogram(col="violet")+facet_grid(~Strata)

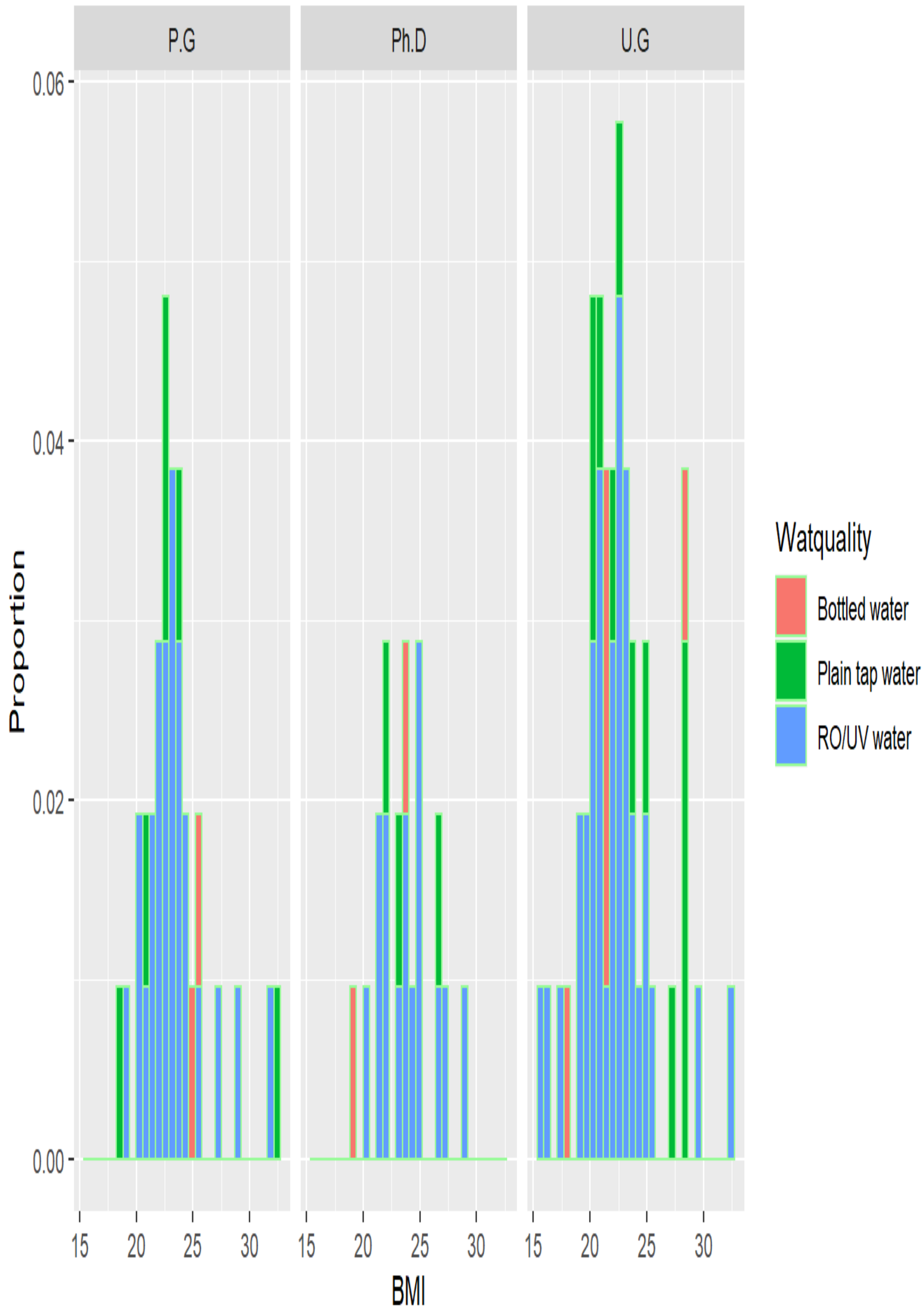
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Question no 4

ggplot(data=students,aes(x=BMI,,y=(..count..)/sum(..count..),fill=Watquality))
+geom_histogram(col="palegreen")+labs(y="Proportion")+facet_grid(~Strata)

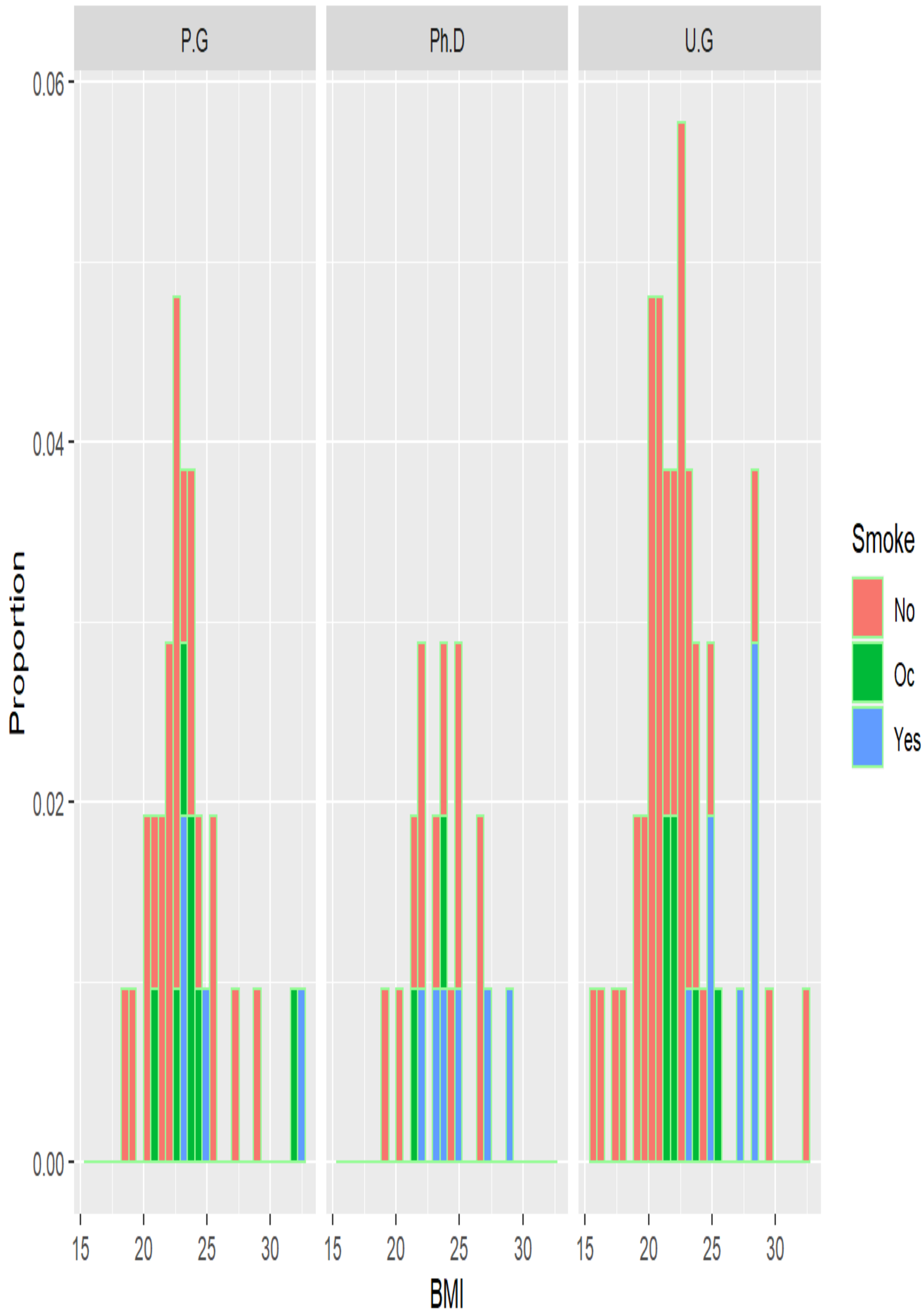
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Question no 5

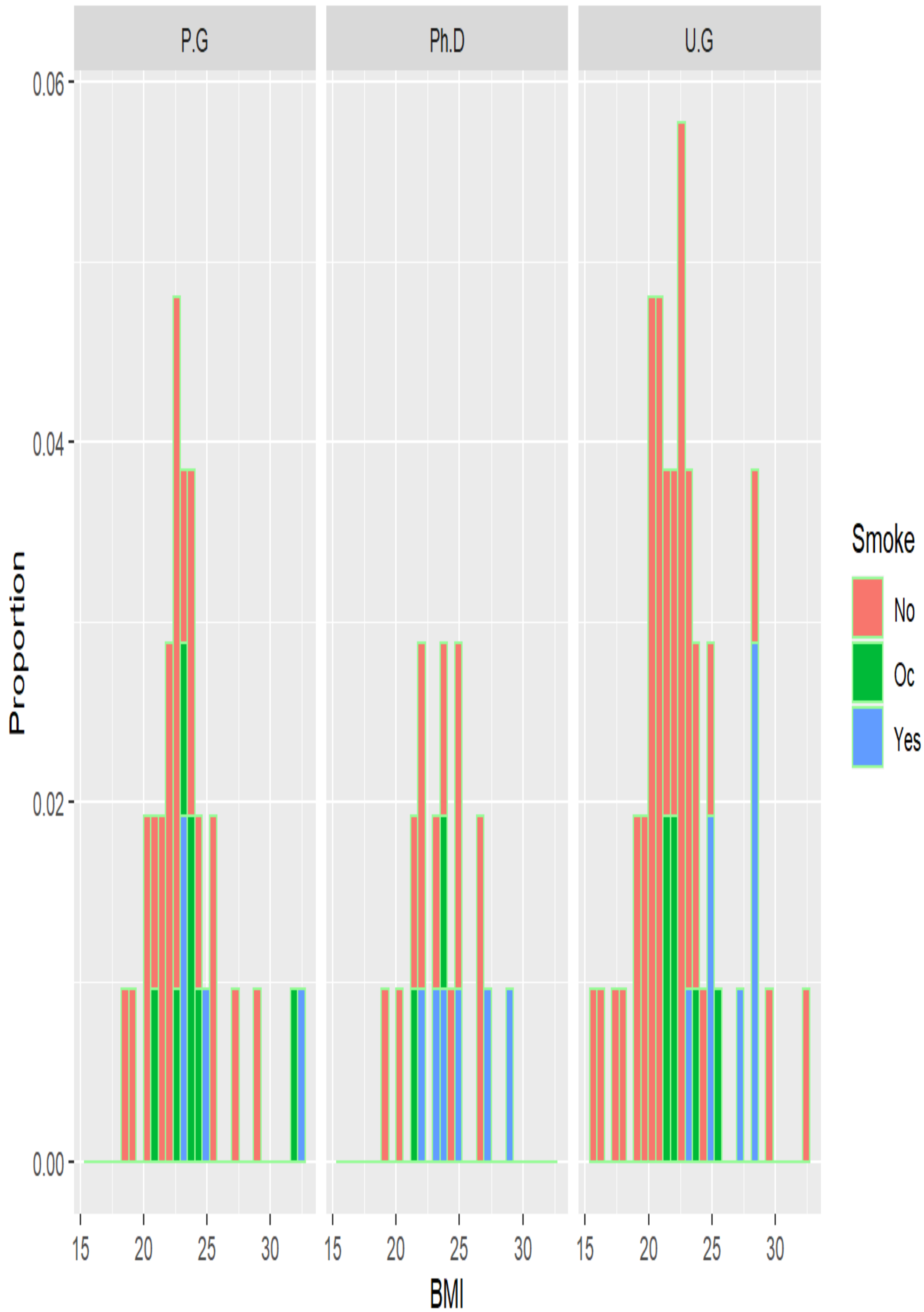
ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=Smoke))
+geom_histogram(col="palegreen")+labs(y="Proportion")+facet_grid(
~Strata)

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




```
ggplot(data=students,aes(levels(students),x=BMI,y=(..count..)/sum(..count..),fill=Smoke))
+geom_histogram(col="palegreen")+labs(y="Proportion")+facet_grid(
~Strata)
```

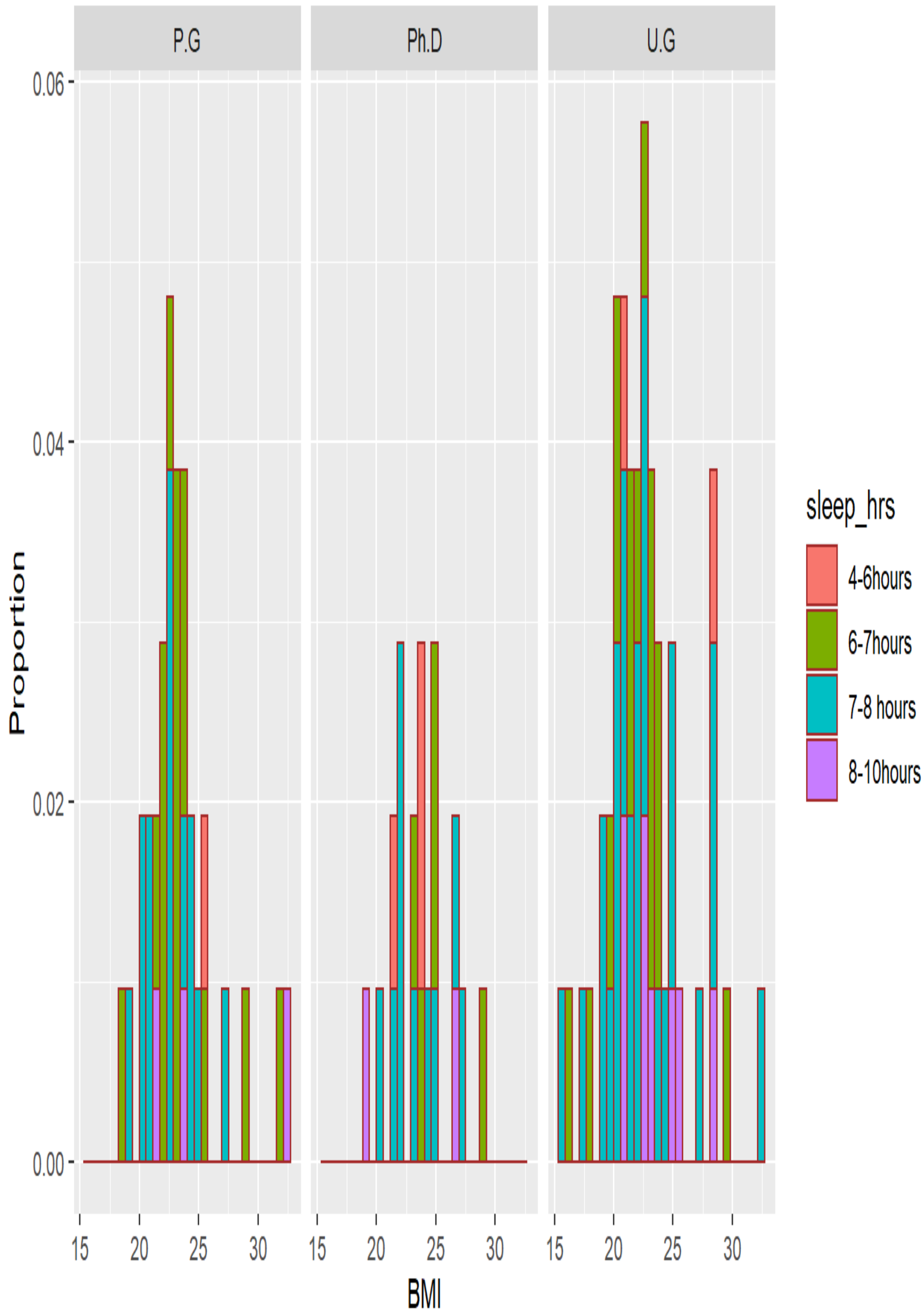
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



#Question no 8

```
ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=sleep_hrs))+geom_histogram(col="brown")
+labs(y="Proportion")+facet_grid(~Strata)
```

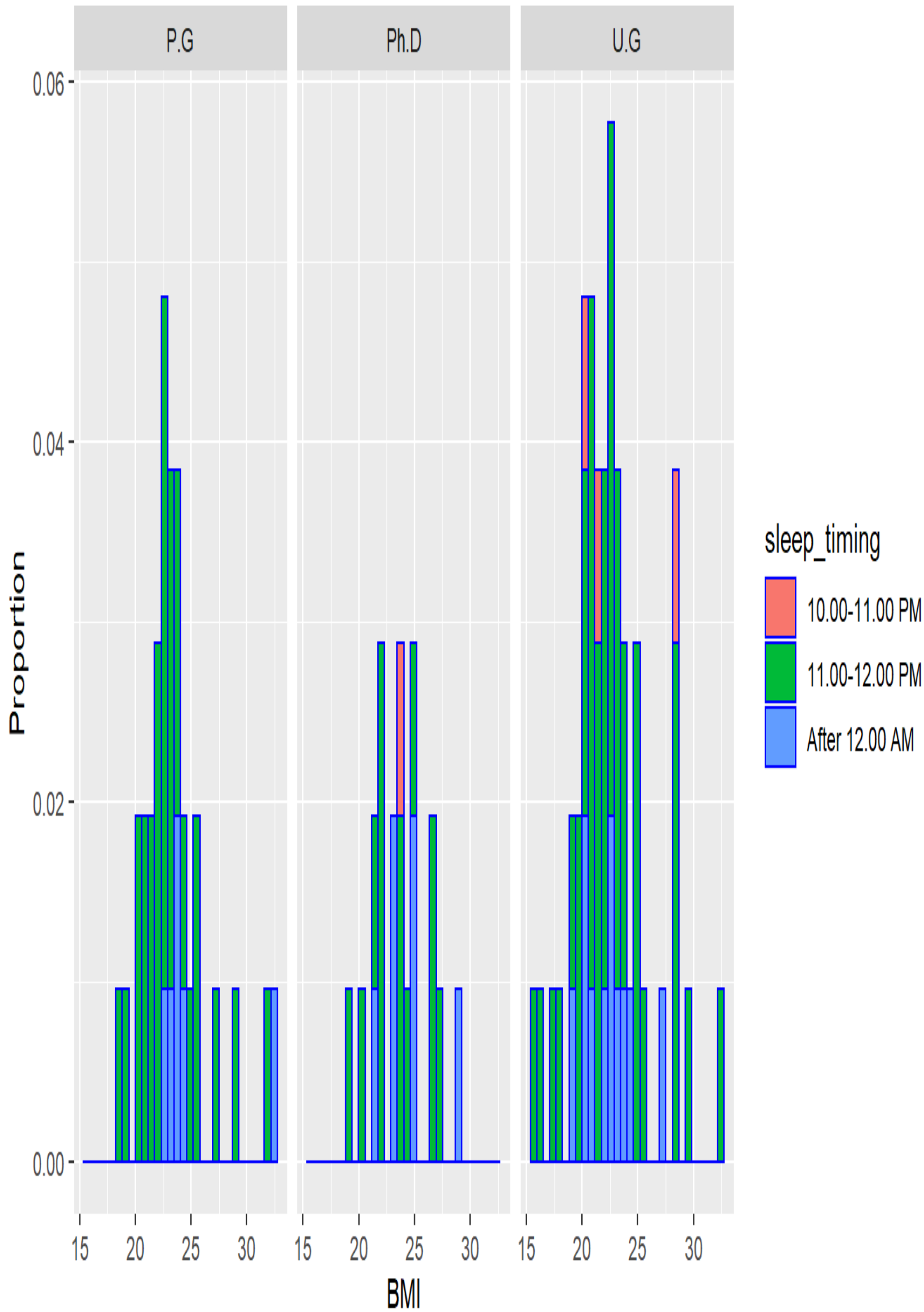
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



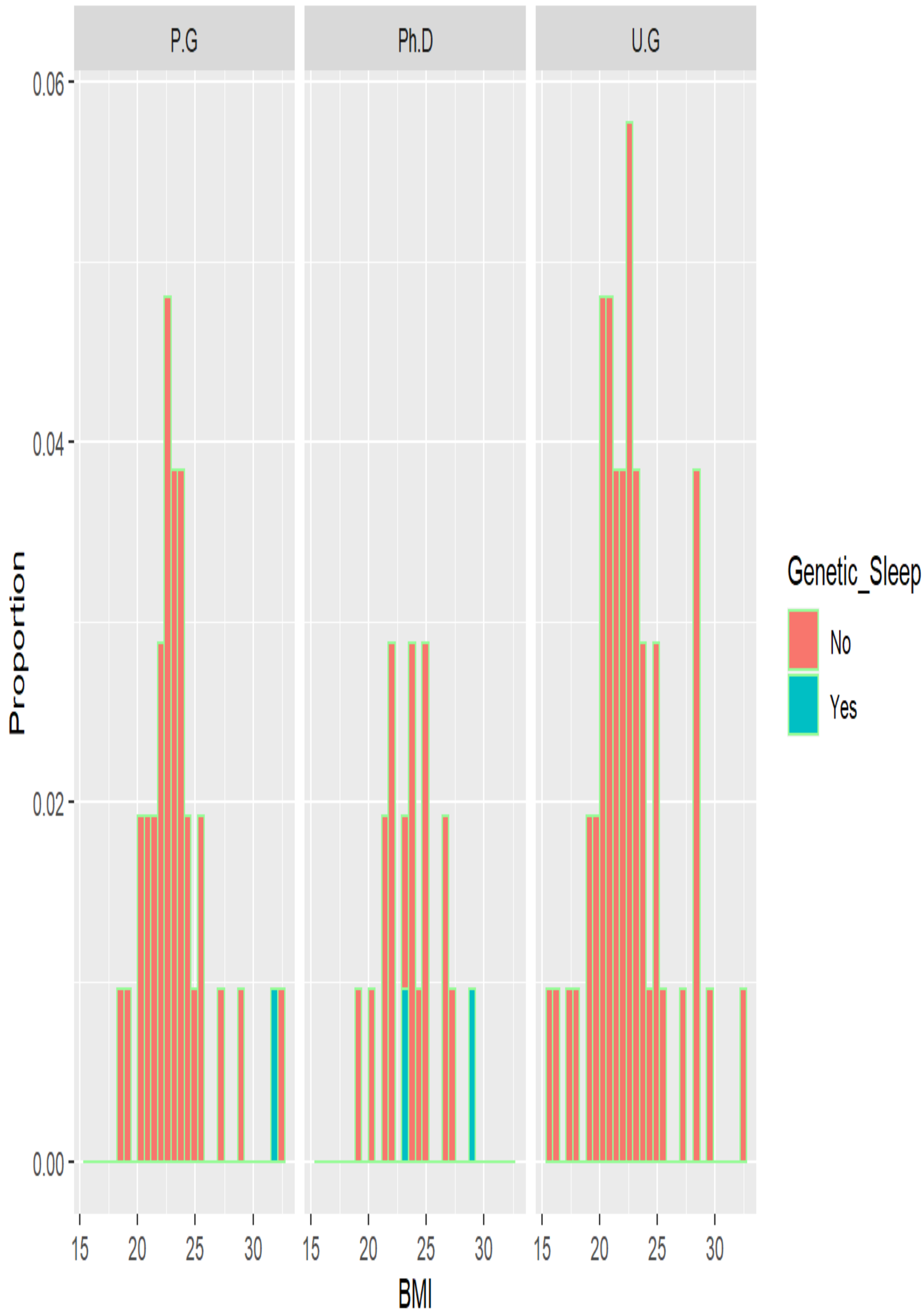
#Question no 9

ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=sleep_timing))
+geom_histogram(col="blue")+facet_grid(~Strata)+labs(y="Proportion")

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



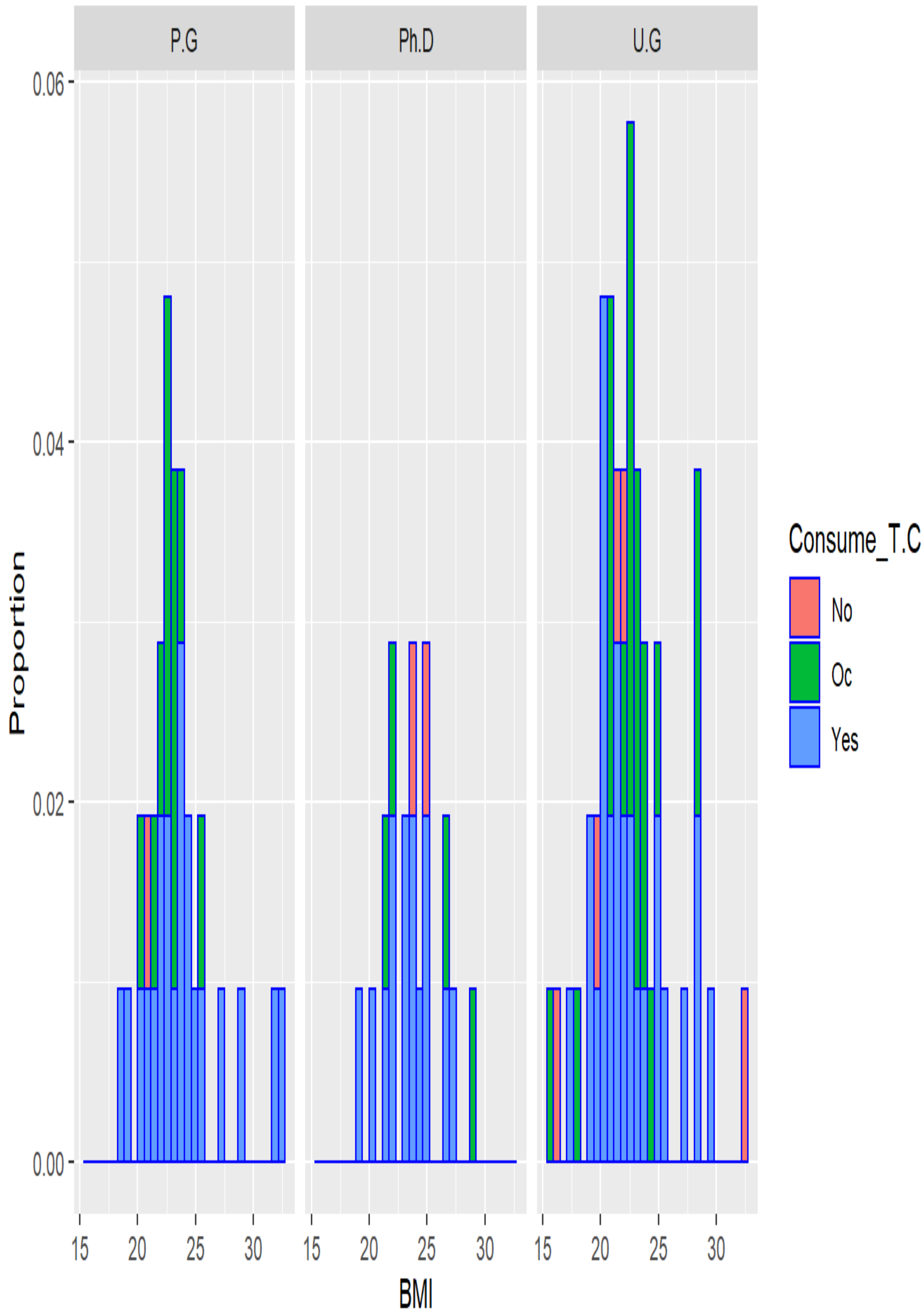

```
#Question no 10
ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=Genetic_Sleep))
+geom_histogram(col="palegreen")+facet_grid(~Strata)+labs(y="Proportion")

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



#Question no 11

```
ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=Consume_T.C))
+geom_histogram(col="blue")+facet_grid(~Strata)+labs(y="Proportion")
```

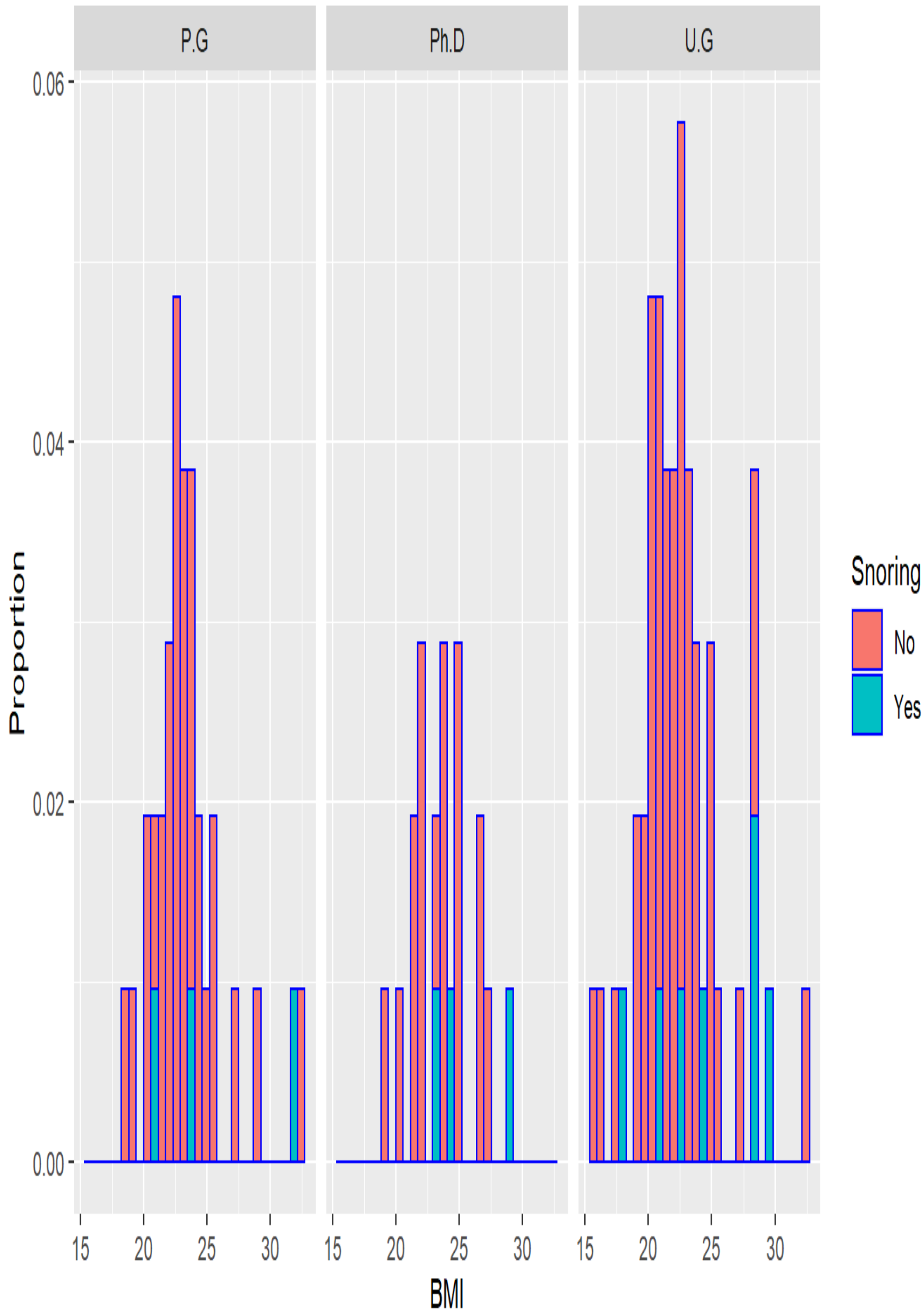
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



#Question no 12

ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=Snoring))+geom_histogram(col="blue")
+facet_grid(~Strata)+labs(y="Proportion")

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

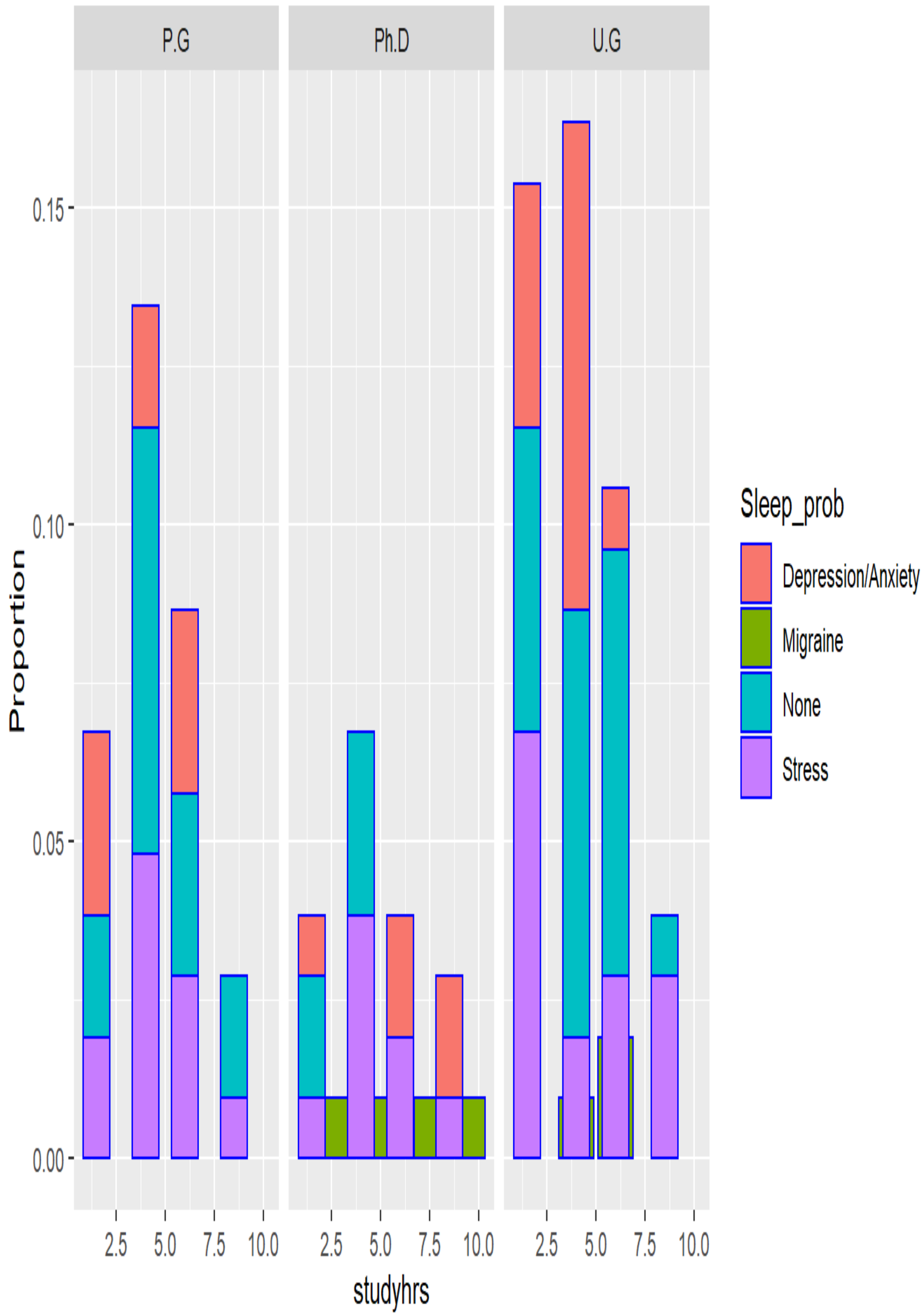


#Question no 14

ggplot(data=students,aes(x=studyhrs,y=(..count..)/sum(..count..),fill=Sleep_prob))+geom_bar(col="blue")
+facet_grid(~Strata)+labs(y="Proportion")

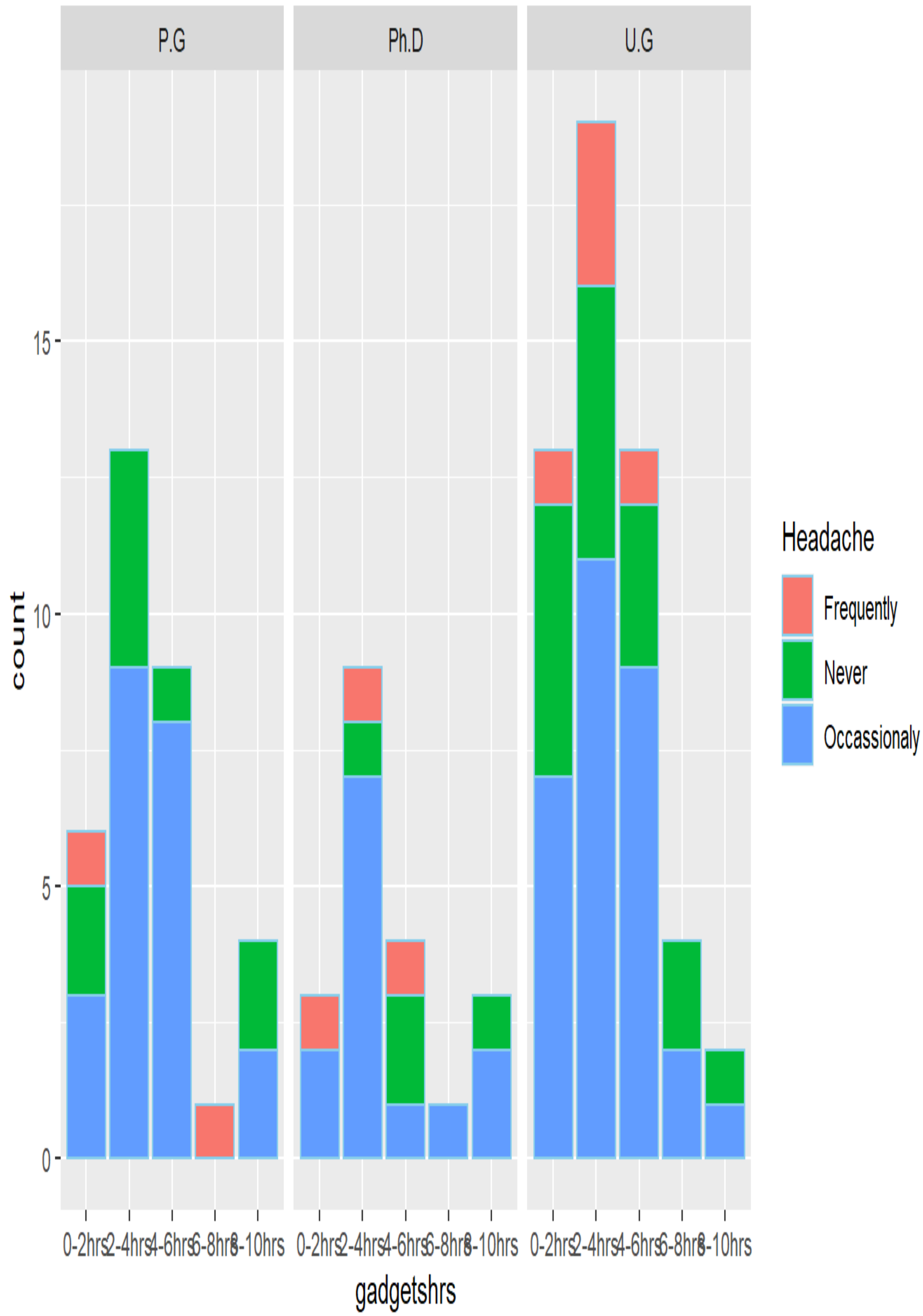
Warning: position_stack requires non-overlapping x intervals

Warning: position_stack requires non-overlapping x intervals



#Question no 16

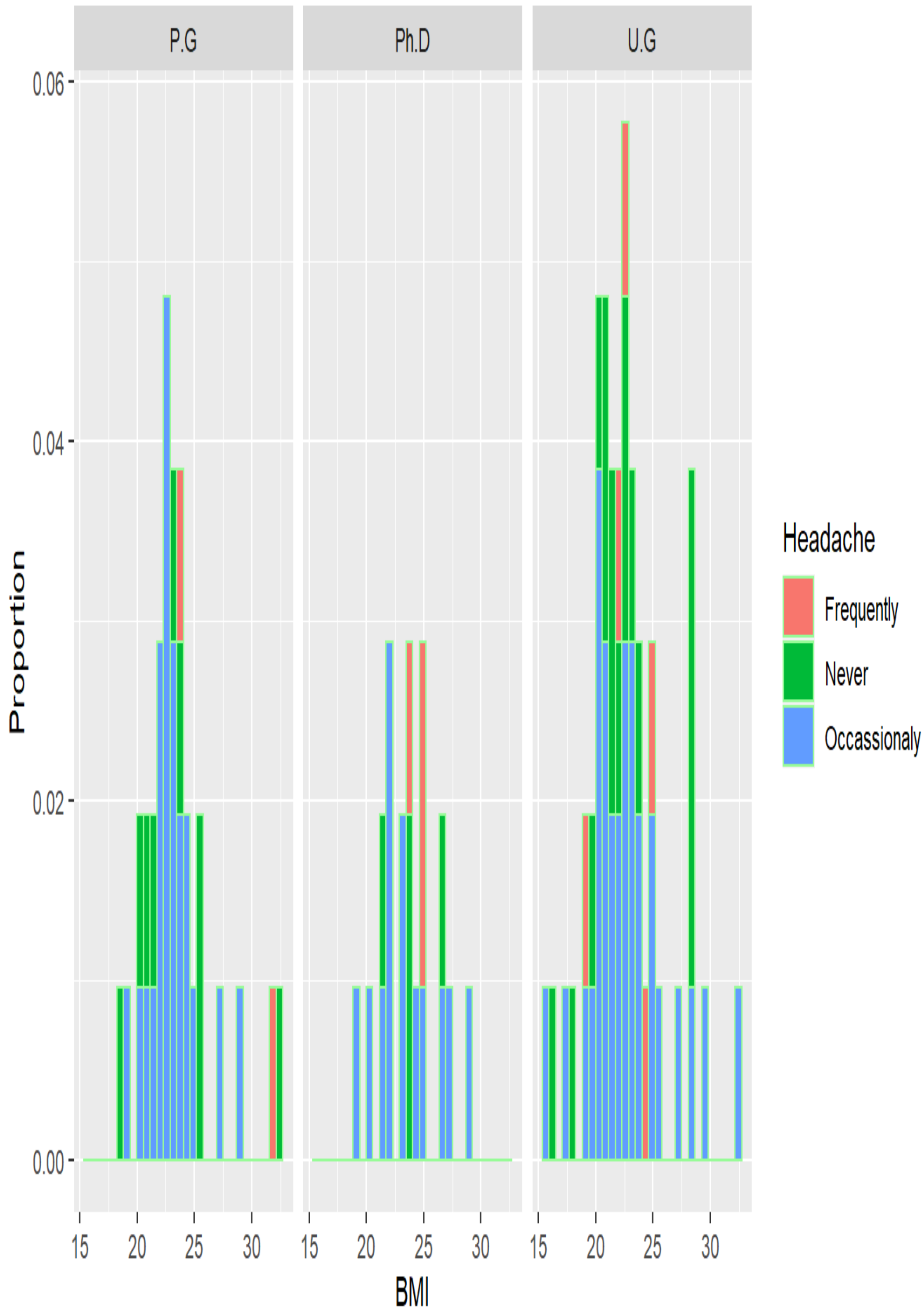
```
ggplot(data=students,aes(x=gadgetshrs,fill=Headache))+geom_bar(col="skyblue")+facet_grid(~Strata)
```



#Question no 17

```
ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=Headache))
+geom_histogram(col="palegreen")+facet_grid(~Strata)+labs(y="Proportion")
```

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

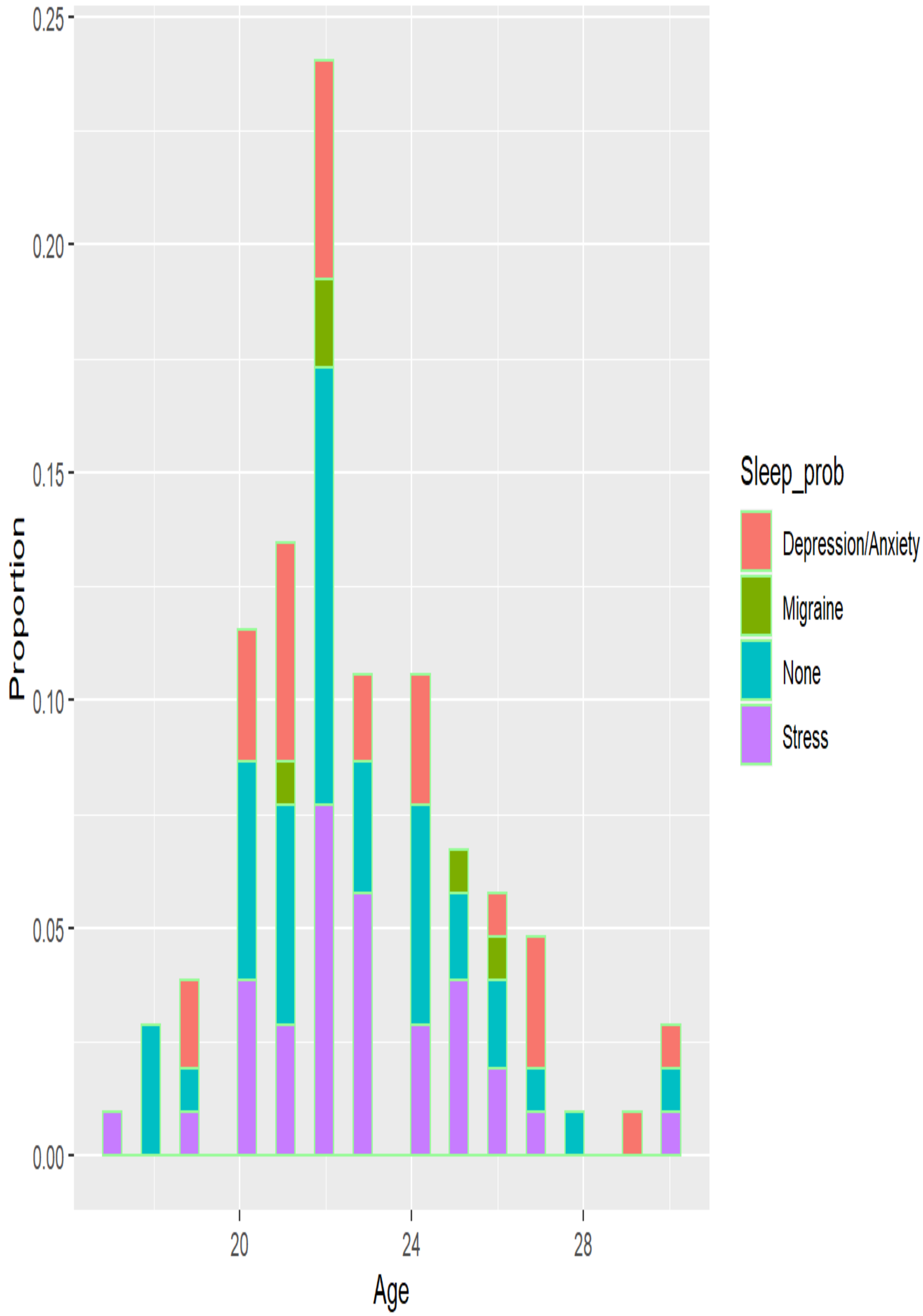


#Question no 18

ggplot(data=students,aes(x=Age,y=(..count..)/sum(..count..),fill=Sleep_prob))

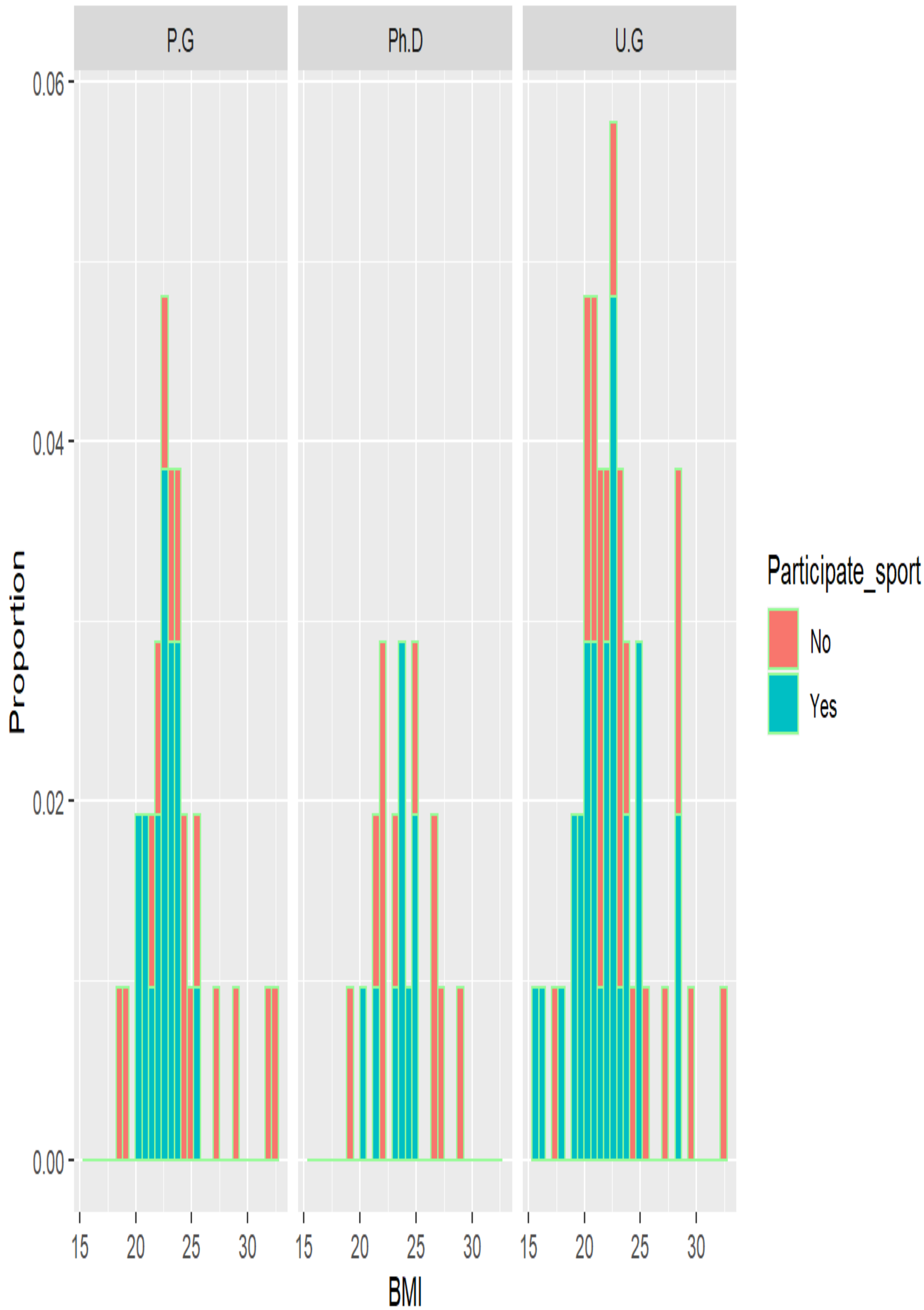
+geom_histogram(col="palegreen")+labs(y="Proportion")

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



#Question no 25

```
ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=Participate_sport))
+geom_histogram(col="palegreen")+labs(y="Proportion")+facet_grid(~Strata)

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

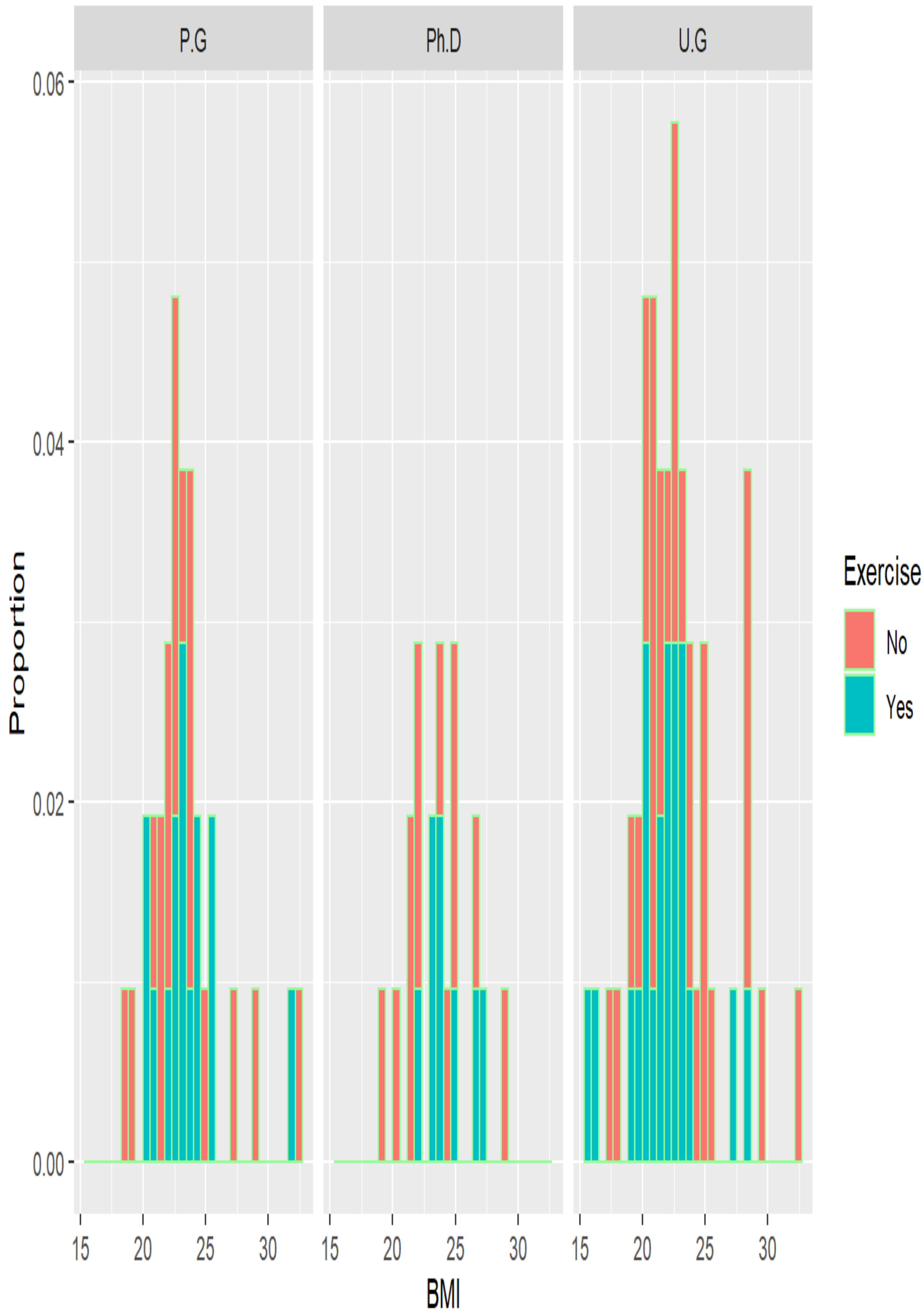


#Question no 26

ggplot(data=students,aes(x=BMI,y=(..count..)/sum(..count..),fill=Exercise))

+geom_histogram(col="palegreen")+labs(y="Proportion")+facet_grid(~Strata)

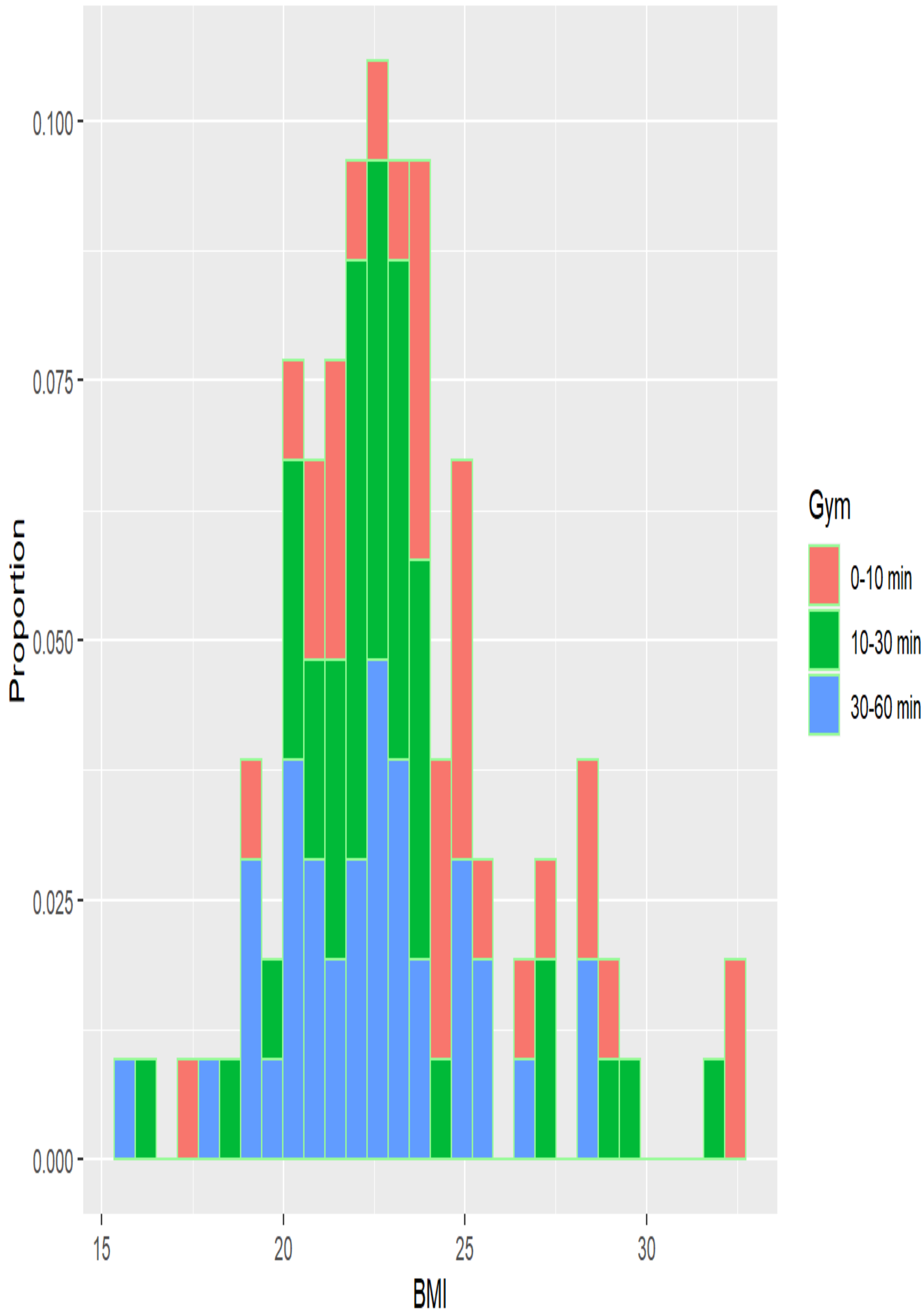
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Question no 27

ggplot(data=students,aes(x=BMI,,y=(..count..)/sum(..count..),fill=Gym))+geom_histogram(col="palegreen") +labs(y="Proportion")

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



t_c=students$Consume_T.C

t_c

```
##   [1] "Oc"  "Oc"  "Oc"  "Yes" "Oc"  "Yes" "Yes" "Yes" "Yes" "Yes" "Oc"  "Oc"
##  [13] "Yes" "No"  "Oc"  "Yes" "Oc"  "Yes" "Yes" "Oc"  "Yes" "Yes" "Yes" "Yes"
##  [25] "Yes" "Yes" "Yes" "No"  "Oc"  "Oc"  "Yes" "Yes" "Oc"  "No"  "Yes" "Yes"
##  [37] "Oc"  "No"  "No"  "Yes" "Yes" "Yes" "Yes" "Oc"  "Oc"  "Yes" "Oc"  "Oc"
##  [49] "Yes" "Oc"  "Oc"  "Oc"  "Yes" "Oc"  "Yes" "Yes" "Oc"  "Yes" "Yes" "Yes"
```

```
## [61] "Yes" "Oc"  "Yes" "Yes" "Yes" "No"  "Yes" "Oc"  "Yes" "Yes" "Oc"  "Yes"
## [73] "Yes" "Yes" "Oc"  "Oc"  "Yes" "Oc"  "Yes" "Yes" "Oc"  "Yes" "Yes" "Oc"
## [85] "Yes" "Yes" "No"  "Oc"  "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes"
## [97] "Yes" "Yes" "Yes" "Yes" "Oc"  "Oc"  "Oc"  "No"
```

```
str(t_c)
```

```
## chr [1:104] "Oc" "Oc" "Oc" "Yes" "Oc" "Yes" "Yes" "Yes" "Yes" "Yes" "Oc" ...
```

```
BMI=students$BMI
```

Testing of Hypothesis

Chi-square test(): We are applying Chi-square test on the question no 17 & 18 in which we want to see, is there any dependency between headache and stress, migraine, Depression anxiety?

Question.17: Do you suffer from headache?

Distribution of Headache in Students

Strata	Never	Occasionally	Frequently	N_A	Total
UG	5	38	5	1	59
PG	10	27	4	0	41
Ph.D	5	17	5	1	28
Total	20	72	14	2	128

Question.18: Have you ever had any of the following?

Stress & Migraine in Students

Strata	Stress	Depression_Anxiety	None	NA_	TOTAl
UG	20	13	23	0	59
PG	12	11	17	0	41
Ph.D	11	5	10	0	28
Total	23	29	50	0	128

\(H_o\): There is no dependency between them.

\(H_1\): There is dependency between headache and stress, migraine Depression anxiety.

At the 5% level of significance.

Strata-wise chi-squared testing

For UG(Undergraduates):

Chi-Squared testing in PG Students

Headache	Stress_	Depression_Anxiety	Migraine_problem	None_	Row_Frequency
Never	1	13	2	10	16
Occasionally	16	11	1	11	38
Frequently	2	5	1	1	5
Column Frequency	19	29	4	22	39

```
M1 = matrix(c(1,16,2,3,10,1,2,1,1,10,11,1),ncol = 4)
dimnames(M1) = list(c("Never","Occassionally","Frequently"),c("Stress","Depression/
Anxiety","Migraine","None"))
M1

##               Stress Depression/Anxiety Migraine None
## Never              1                  3        2   10
## Occassionally     16                 10        1   11
## Frequently         2                  1        1    1

chisq.test(M1)

## Warning in chisq.test(M1): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  M1
## X-squared = 11.764, df = 6, p-value = 0.06744
```

Conclusion:- Since p-value(0.06744) is greater than level of significance(0.05), we have enough evidence to fail to reject the null hypothesis. Therefore we may conclude that there is no dependency among headache, stress, migraine and depression.

For PG(Postgraduates):-

Chi-Squared testing in Ph.D Students

Headache_PG	Stress_PG	Depression_anxiety_PG	Migraine_problem_PG	None_PG	Row_Frequency_PG
Never	2	3	0	5	10
Occasionally	9	6	0	11	26

Headache_PG	Stress_PG	Depression_anxiety_PG	Migraine_problem_PG	None_PG	Row_Frequency_PG
Frequently	1	2	1	0	4
Column Frequency	12	11	1	16	40

M2=matrix(c(2,9,1,3,6,2,0,0,1,5,11,0),ncol=4)
dimnames(M2)=list(c("Never","Occasionally","Frequently"),c("Stress","Depression/Anxiety","Migrane","None"))
M2

```
##              Stress Depression/Anxiety Migrane None
## Never             2                  3       0    5
## Occasionally      9                  6       0   11
## Frequently        1                  2       1    0
```

chisq.test(M2)

```
## Warning in chisq.test(M2): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  M2
## X-squared = 12.38, df = 6, p-value = 0.05401
```

Conclusion:- Since p-value(0.05401) is greater than level of significance(0.05) , we have enough evidence to accept the null hypothesis . Therefore we may conclude that there is no dependency among headache , stress , migrane and depression.

For Ph.D:-

Chi-Squared testing in Ph.D Students

Headache_PG	Stress_PG	Depression_anxiety_PG	Migraine_problem_PG	None_PG	Row_Frequency_PG
Never	2	3	0	5	10
Occasionally	9	6	0	11	26
Frequently	1	2	1	0	4
Column Frequency	12	11	1	16	40

M3=matrix(c(0,7,4,0,5,0,1,1,0,4,4,1),ncol=4)

```
dimnames(M3)=list(c("Never","Occasionally","Frequently"),c("Stress","Depression/
Anxiety","Migrane","None"))
M3

##              Stress Depression/Anxiety Migrane None
## Never             0                  0       1    4
## Occasionally      7                  5       1    4
## Frequently        4                  0       0    1

chisq.test(M3)

## Warning in chisq.test(M3): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  M3
## X-squared = 12.388, df = 6, p-value = 0.05385
```

Conclusion:- Since p-value(0.05385) is greater than level of significance(0.05) , we have enough evidence to accept the null hypothesis . Therefore, we may conclude that there is no dependency among headache, stress, migraine and depression.

For Full Sample:-

```
Headache = c('Never','Occasionally','Frequently','Column Frequency')
Stress_FS = c(3,32,7,42)
Depression_anxiety_FS = c(6,21,3,30)
Migraine_problem_FS = c(3,2,2,7)
None_FS = c(19,26,2,47)
Row_Frequency_FS = c(31,81,14,126)
df_full =
data.frame(Headache,Stress_FS,Depression_anxiety_FS,Migraine_problem_FS,None_FS,Row_Frequency_F
S)
kable(df_full,caption = "Chi-Squared testing of total Students")%>%kable_styling(bootstrap_options =
"striped",font_size = 14)
```

Chi-Squared testing of total Students

Headache	Stress_FS	Depression_anxiety_FS	Migraine_problem_FS	None_FS	Row_Frequency_FS
Never	3	6	3	19	31

Headache	Stress_FS	Depression_anxiety_FS	Migraine_problem_FS	None_FS	Row_Frequency_FS
Occasionally	32	21	2	26	81
Frequently	7	3	2	2	14
Column Frequency	42	30	7	47	126

```
M4=matrix(c(3,32,7,6,21,3,3,2,2,19,26,2),ncol=4)
dimnames(M4)=list(c("Never","Occasionally","Frequently"),c("Stress","Depression/
Anxiety","Migrane","None"))
M4
```

```
##              Stress Depression/Anxiety Migrane None
## Never             3                  6       3   19
## Occasionally     32                 21       2   26
## Frequently        7                  3       2    2
```

```
chisq.test(M4)
```

```
## Warning in chisq.test(M4): Chi-squared approximation may be incorrect
```

```
## 
##  Pearson's Chi-squared test
## 
## data:  M4
## X-squared = 19.357, df = 6, p-value = 0.003601
```

Conclusion:- Since p-value(0.003601) is less than level of significance(0.05) , we have enough evidence to reject the null hypothesis . Therefore we may conclude that there is dependency among headache , stress , migraine and depression.

Z-test

We want to compare BMI of our population to the BMI of the sample to elicit the information that most of the students lie in which category, either they are healthy category or underweight or overweight

Testing Hypothesis

\(H_0: p = 0.5\) (around 50% populatoon of Habib Hall lie in the healthy category of BMI)

\(H_1: p ≠ 0.5\) (there will be more or less students who lie in the healthy category of BMI)

```
library(BSDA)
```

```
## Warning: package 'BSDA' was built under R version 4.0.5
```

```
## Loading required package: lattice

##
## Attaching package: 'BSDA'

## The following object is masked _by_ '.GlobalEnv':
##
##     Stress

## The following object is masked from 'package:datasets':
##
##     Orange
```

```
bmi = students$BMI
n = length(bmi)
#intoducing an empty list
len <- 0
x <- vector(mode = "list", length = len)
x
```

```
## list()
```

```
class(x)
```

```
## [1] "list"
```

```
#Searching BMI who comes under in Healthy Category
for (i in bmi){
  if(i > 18.5 && i <= 24.9){
    x = append(x, list(i))
  }
}
X = length(x)
X
```

```
## [1] 77
```

```
binom.test(X, n, p = 0.5, alternative = "two.sided", conf.level = 0.95 )
```

```
##
```

```
##  Exact binomial test
##
## data:  X and n
## number of successes = 77, number of trials = 104, p-value = 9.702e-07
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.6451899 0.8214254
## sample estimates:
## probability of success
##              0.7403846
```

Conclusion Since, p-value is less than \(\alpha = 0.05\) we reject our null hypothesis it means more than half of the total students are either underweight or overweight.

RESULTS AND CONCLUSIONS

- Most of the students take Dining food & extra diet are much more than those who take Tiffin service & self cooked food combined.
- Most students in the hall consumes 2-3 litres of water daily which is lethan the recommended by Doctors.
- Most of the students in each strata and overall prefers Fruits and energy drink in the form of extra diet. Although some of the students also like to have Junk food and Dry fruits.
- The proportion of the population of Habib hall who consume RO/UV water falls under the healthy category of BMI.
- In UG category, most students are those who never smoke yet they come under healthy category of BMI. Also in PG & P.hD, many students are those who smoke regularly and occasionaly come under healthy range of BMI.
- Students whose height is in between 160-180 cm comes under the healthy catrgory of BMI(18.5 to 24.9).
- As the age of the students increases, BMI also increses.
- students who sleeps 6-8 hours in a day are healthy.
- Most of the students who sleep between the interval 11.00 PM-12.00 AM and after 12.00 AM comes under the healthy category of BMI.
- Approximately 97% of the total students consisting of UG, PG & Ph.D does not have any genetic sleep problem.
- Most of the students in each strata who consume Tea & Coffee(regularly & Occasionaly) comes under the healthy category of BMI. It means that consumption of Tea & Coffee does not effect BMI.
- 50.84%, 56.09% & 57.14% of UG, PG and Ph.D respectively suffers stress. Around 2% of students have consistent sleep.
- On comparing study hours to Stress, Migraine & Depression, 13% of UG population who study less

than 2.5 hours have 50% stress and 20% depression, which is very high proportion in comparison to PG & Ph.D.

- Most students prefer to study Morning time and Late night. Those students who like to study in Day time are very less in number in every category.
- Large number of students spend more than 6 hours on electronic gadgets. However, students from UG category spend most among them.
- About half of the total students suffers headache occassionaly. approx. 11% of the students face frequent headache problems.
- Students who are at the age of 22 are highly exposed to stress and depression problem. Migrane problem is found to be very rare in the lower and upper age group.
- More than half of the students combining UG, PG & Ph.D does not agree with the fact that the health & hygiene condition are good at their residential hall.
- More than 50% of the students are satisfied with treamtent facilities and rest of the students are not.
- Students agreed on the fact that the sanitation & hygiene level of Habib Hall is very vicious or poor.
- More than 50% of the total students accept that the environment of their surrounding is Normal.
- Half of the total students are involved in sports activities. Moreover, those students who are involved in sports activities, their BMI comes under the healthy category.
- Larger proportion of the students who indulge in physical activities atleast 10-30 minutes, comes under the healthy category of BMI.
- There is no dependency between headache and stress, migraine & depression/anxiety in each stratum but when it comes to the whole sample there is the dependency exists between headache and stress, migraine & depression/anxiety.
- When we applied z-test on our sampling frame we have found that more than half of the total students are either underweight or overweight.