* Our project is to analyze the sleep patterns of IITH students
* Its **objective** is to identify various factors which affect our sleep and the way they affect us.
* The **variables of interest** for this study are

**Response variables** (Response Variable is the focus of question in a study)

1) Number of hours of sleep(night)(**continuous** numerical/**Ordinal** Categorical)

2) start time of sleep (night)(**continuous** numerical)

**Explanatory variables** (variable(we know) that affect response variables)

1) Whether a person consume caffeine, yes/no (**Nominal/Regular** Categorical)

2) How many times caffeinated drinks are consumed per day(**Ordinal**Categorical)

3) How and at when do people watch lectures (**Nominal/Regular** Categorical)

4) How and when do people submit assignments?(**Nominal/Regular** Categorical)

5) screen time to study(**Ordinal** Categorical / **continuous** numerical)

6) start time of study (**continuous** numerical)

7) screen time other than study(**Ordinal** Categorical / **continuous** numerical)

\*8) Time spent on physical activities(**Ordinal** Categorical)

* As our interest is to analyze natural sleeping patterns we perform an observational study.(we assume the effect of any **confounding** (extraneous variables that misleads us to make a false relation between response and explanatory variables) variables to be negligible as we are concentrating on sleeping patterns of student life)
* We used sampling more specifically **volunteer sampling** for collection of data by sending mail to every student at IITH but only a few of them volunteered to respond. To make as many people respond as possible we also send remainders. And managed to collect data from 132 people, the data is diverse with data from all years of UG, PG and Phd students.

STUDY ON NUMBER OF HOURS OF SLEEP(single continuous **numerical** variable – using **box plot**)

* From the central tendency calculation , we can infer that the population of IITH on an average sleeps more than 6 hours.
* We tested this inference using **hypothesis testing using t-distribution.**
* With a test statistic of 5.84 and critical value of 2.35 at a 1.0% level of significance, we have enough statistical evidence to reject the null hypothesis.
* So, there is enough evidence to conclude that IITH students are getting more than 6hrs of sleep a day.
* We also plotted the **sampling mean distribution** for the number of hours of sleep by selecting random samples of varying sizes greater than 30. We can observe that as the random sample size increases the distribution is tending to be normal. From this we can observe that our data is **not highly skewed** and good representative of the population.

SLEEP TIME (single continuous **numerical** variable – using **histogram**)

* For calculating central tendency for sleep we first normalized the data by **shifting origin**. If the time value is directly taken in a 24 hour clock then the mean calculated will be wrong because of the jump in value of time from 12 pm(24 hours) to 1 am(1 hours) .
* ex: if we have a sleep time of 1 am (01 hours) and 11pm (23 hours) mean we get 12 hours ( 12 noon) which is false.
* For our data we took the origin as 8 pm.
* From the histogram we can infer that the data is **slightly right skewed and unimodal.**
* We also plotted the **sampling mean distribution** for sleep time by selecting random samples of varying sizes greater than 30. We can observe that as the random sample size increases the distribution is tending to be normal. From this we can observe that our data is **not highly skewed** and good representative of the population.
* We can say that with **95% confidence** that the population mean time at which a person sleeps lies between (1:07 hours, 2:09 hours).

STUDY TIME (single continuous **numerical** variable – using **histogram**) effect on number of hours of sleep

* We used the same technique we used in Sleep time to normalize the data by shifting origin. For study time origin is taken as 5 AM.
* Using this, central tendencies are calculated and histogram is plotted.
* From the histogram we can infer that the study time is near symmetric and **bimodal**.
* From the **side by side boxplot between** study time(**numerical** variable) and Number hours of sleep(**categorical** variable) we can infer the following
* **Shape:** slightly right skewed, so most people tent to start studying early
* **Center:** people with moderate sleep tend to start studying earlier than people with extreme sleep habits.
* **Spread:** The IQR of people who sleep more is less, so they consistently start studying in the evenings.
* We also plotted the **sampling mean distribution** for study time by selecting random samples of varying sizes greater than 30. We can observe that as the random sample size increases the distribution is tending to be normal. From this we can observe that our data is **not highly skewed** and good representative of the population.
* We can say that with 95% confidence that the population mean time at which a person studies lies between (15:12 hours,17:06 hours).

**Effect of caffeine**

Relation between number of Caffeinated drink servings, Number of hours of sleep

* We ploted a side by side boxplot between the number of servings of caffeinated drinks(**Ordinal** Categorical) and Number hours of sleep(**categorical** variable).
* From the box plot we can infer the following:
* **Shape:** for people who dont consume caffeine the data is right skewed so they tend to sleep less but this value is greater than the number of sleep hours of other categories which consume caffeine.
* **Center:** All categories have nearly the same mean.
* **Spread:** We can observe that IQR decreases as caffeine consumption increases, so people who consume caffeine consistently sleep for the same amount of time.
* **Outliers:** There are a considerable number of outliers among the moderate caffeine consuming categories.
* We also calculated **95% confidence intervals** for the average number of hours of sleep in each caffeine consuming category, using **t-distribution** of sample mean.(values can be found in the report).
* We conducted **Hypothesis testing** on Hypothesis “people who take caffeine sleep less when compared with people who dont.”
* test statistic is 0.9180487079966966
* t is 1.661961083996941
* The evidence is insufficient to conclude that those who take caffeine sleep less than those who dont take caffeine.

Relation between Caffeine intake, Sleep Times

* We ploted a side by side boxplot between the number of servings of caffeinated drinks(**Ordinal** Categorical) and Sleep Times(**numerical** variable).
* From the box plot we can infer the following:
* **Shape:** The box plot for high caffeine consumers is right skewed. More people in this category sleep early but this is late when compared with other categories.
* **Center:** All categories have nearly the same mean.
* **Spread:** We can observe that **IQR** of high caffeine consumers is more when compared with other categories, so they **donot sleep consistently** at the same time.
* **Outliers:** There are a few outliers among less or no caffeine consuming categories.

Effect of academic work habits on studies

Relation between lecture watching patterns and number of hours of sleep

* We plotted a side by side boxplot between the number of hours of sleep(**numerical** variable) and lecture watching patterns(**categorical** variable)
* From the box plot we can infer the following:
* **Shape:** Only category with left skewness is people who watch only live lectures, so they tend to sleep for more time among themselves, which is different when compared with other categories.
* **Center:** The average amount of sleep for the category of people watching recordings regularly is slightly higher than other categories.
* **Spread:** The **IQR** of the category of people who watch both live lectures and recordings is the least, so they sleep more consistently when compared with other categories. Also, the **range** of sleep hours for the category of people who binge watch lectures is very high when compared with others, so the number of hours they sleep is very unpredictable.
* We also plotted the segmented bar chart with lecture watching patterns(**categorical** variable), Number hours of sleep(**categorical** variable) with number of people, for better understanding of the data.

Relation between Assignment Submission Patterns and number of hours of sleep

* We plotted a side by side boxplot between the number of hours of sleep(**numerical** variable) and Assignment Submission Patterns(**categorical** variable)
* From the box plot we can infer the following:
* **Shape:** Only category with right skewness is people who submit the assignment last minute, so they tend to sleep for less time among themselves.
* **Center:** The average amount of sleep for the category of people is slightly increasing based on how close to the deadline they are submitting.
* **Spread:** The **IQR** of the category of people whothe assignment last minute is highest, so they sleep less consistently when compared with other categories.
* We also plotted the segmented bar chart with Assignment Submission Patterns(**categorical** variable), Number hours of sleep(**categorical** variable) with number of people, from which we can infer that more people submit on the last day of deadline.

**EFFECT OF Sports and Entertainment ON AMOUNT OF SLEEP**

Relation between Recreational activities and number of hours of sleep

* We plotted a side by side boxplot between the number of hours of sleep(**numerical** variable) and Recreational activities(**categorical** variable)
* From the box plot we can infer the following:
* **Shape:** Only category with right skewness is people who spend 4-5 hours on recreation, so they tend to sleep for less time among themselves.
* **Center:** The median of all the categories are nearly the same.
* **Spread:** The **IQR** of the category of people who spend 4-5 hours on recreation is highest, so they sleep less consistently when compared with other categories.
* We also calculated **95% confidence intervals** for the average number of hours of sleep in each category, using **t-distribution** of sample mean.(values can be found in the report).
* We conducted **Hypothesis testing** on Hypothesis “A student that spends 1-3 hours on recreational activities a day on average sleeps more than a student that spends more than 4 hours on the same activities a day.”
* test statistic is −2.525969
* t is 1.676
* We can conclude that a student that spends 1-3 hours on recreational activities a day on average sleeps less than a student that spends more than 4 hours on the same activities a day.

Relation between Sports activities and number of hours of sleep

* We plotted a side by side boxplot between the number of hours of sleep(**numerical** variable) and Sports activities(**categorical** variable)
* From the box plot we can infer the following:
* **Shape:** Only category with left skewness is people who spend 4 or more hours on recreation, so they tend to sleep for less time among themselves and exactly the opposite applies for the remaining categories.
* **Center:** The median of people who do not exercise is slightly higher than other categories. The people who do not play sports tend to sleep more than others.
* **Spread:** The **IQR** of the category of people who spend 4-5 hours on recreation is least, so they sleep more consistently when compared with other categories.
* We also calculated **95% confidence intervals** for the average number of hours of sleep in each category, using **t-distribution** of sample mean.(values can be found in the report).
* We conducted **Hypothesis testing** on Hypothesis “A student who does not exercise sleeps less on average than a student who exercises.”
* test statistic is −1.399482137
* t is 1.682
* We can conclude that a student who does not exercise sleeps less on average than a student who exercises.