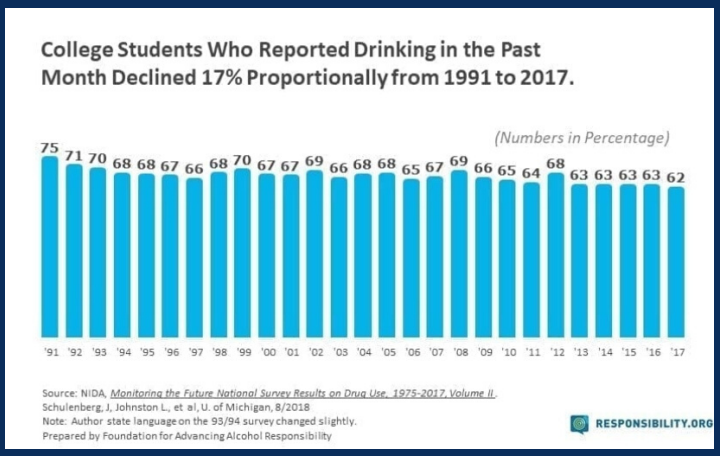
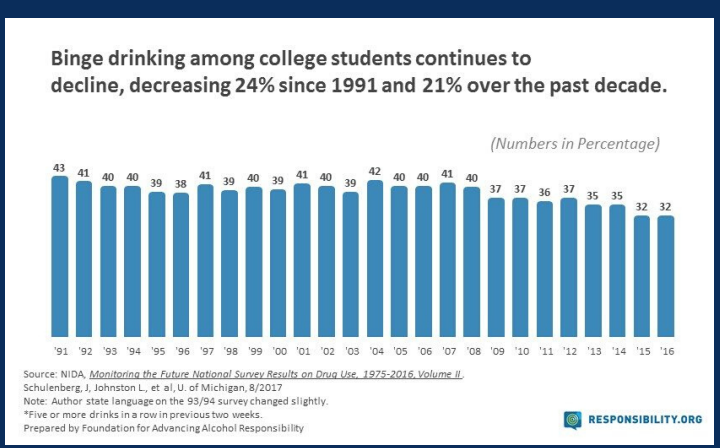
Student Drinking

NANDAN PATEL

WHEN YOUNG STUDENTS IMAGINE THEIR LATER YEARS OF HIGH SCHOOL AND COLLEGE, THEY OFTEN THINK ABOUT THEIR SOCIAL LIVES AND HAVE THIS PICTURE IN THEIR MINDS:

HOWEVER, MOST SCHOOLS ARE NOT FOND OF THEIR STUDENT BODY BEING LABELED AS HEAVY DRINKERS. IN FACT, MOST SCHOOLS, GOVERNMENTS, AND ANTI-DRINKING CAMPAIGNS HAVE TRIED THEIR BEST TO GET RID OF THIS IMAGE. THESE ORGANIZATIONS WANT TO PROVE THAT STUDENTS DO NOT DRINK AS MUCH PEOPLE THINK THEY DO. THEY *TRY* TO MAKE COMPELLING VISUALIZATIONS AND THIS IS THE RESULT:





AWFUL DATA VISUALIZATIONS!

GRAPH 1:

- WHERE IS THE Y-AXIS?
- WHAT DO THE NUMBERS ON TOP OF EACH COLUMN MEAN?
- WHERE IS THE 17% DECLINE HAPPENING?
- WHY IS THE DIFFERENCE BETWEEN THE HIGHEST NUMBER (75) AND LOWEST NUMBER (62) SO SMALL DESPITE THERE BEING A DIFFERENCE OF 13?
- WHAT DOES MONTH HAVE TO DO WITH THIS?

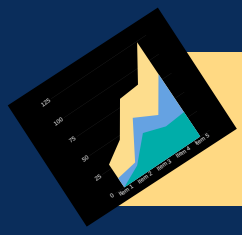
GRAPH 2:

- ALL OF THE ISSUES OF GRAPH 1
- WHAT DEFINES BINGE DRINKING?

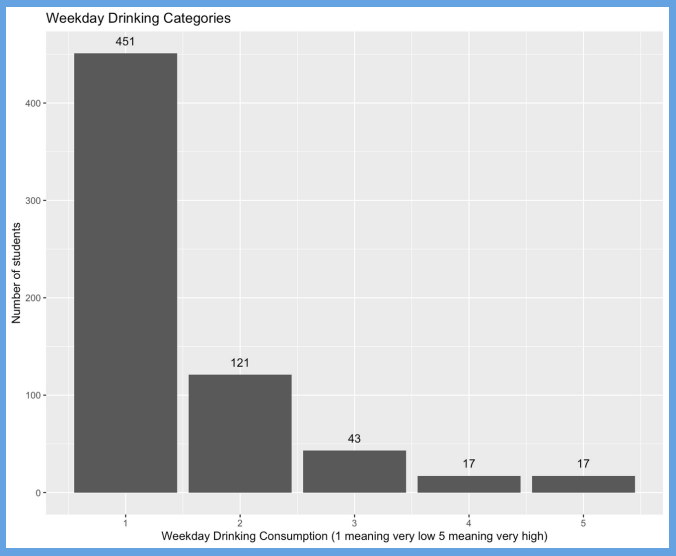
ALONGSIDE THESE HORRIFIC GRAPHS, SCHOOLS PUT OUT STATEMENTS THAT TELL INCOMING STUDENTS THAT THEIR CURRENT STUDENTS ARE NOT HEAVY DRINKERS. AS A PART OF ONBOARDING, RUTGERS TELLS ITS STUDENTS THAT **"2 OUT OF 3 RUTGERS STUDENTS STOP AT THREE DRINKS OR FEWER, WHILE 1 IN 5 DON'T DRINK AT ALL".** THAT IS GREAT AND ALL BUT MOST STUDENTS BELIEVE THAT SCHOOLS MUST MAKE THESE STATEMENTS IN ORDER TO CURB DRINKING. IN OTHER WORDS, THEY SAY IT BECAUSE THEY HAVE TO. STUDENTS THINK THEY COULD BE CHERRY PICKING DATA. THE MAIN ISSUE BECOMES:

Who Do We Trust ?

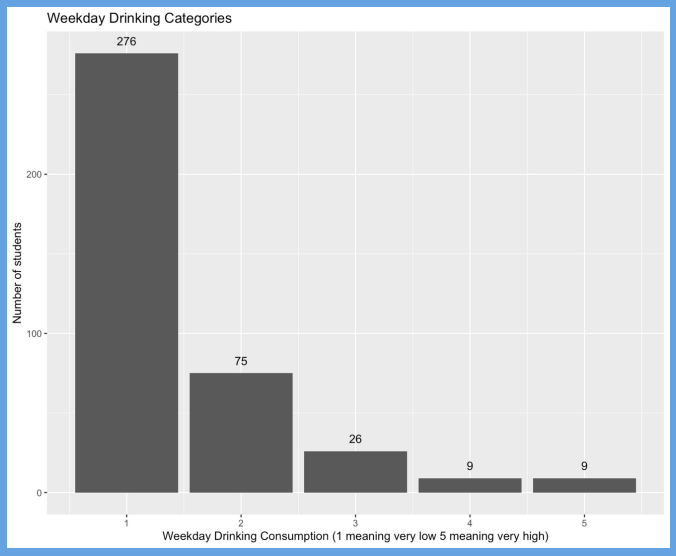
I DECIDED TO TAKE A CLOSER LOOK AT HIGH SCHOOL STUDENT ALCOHOL CONSUMPTION DATA. I WANTED TO LEARN IF STUDENTS ARE ACTUALLY HEAVY DRINKERS, IF THERE ARE REPERCUSSIONS TO DRINKING A LOT, AND WHICH STUDENTS ARE MORE LIKELY TO DRINK. THE DATA COMES FROM UCI MACHINE LEARNING / KAGGLE AND HAS 2 DATA SETS LOOKING AT STUDENTS STUDYING PORTUGUESE AND MATH IN PORTUGAL.



BASIC GRAPHS

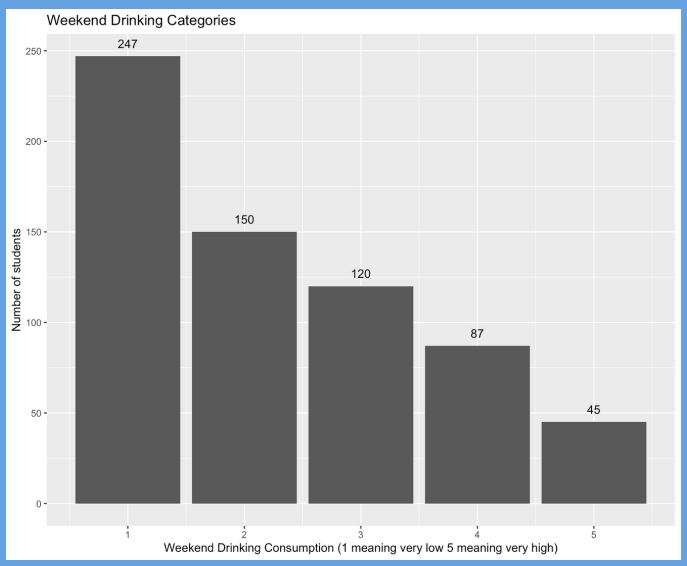


PORTUGUESE STUDENTS

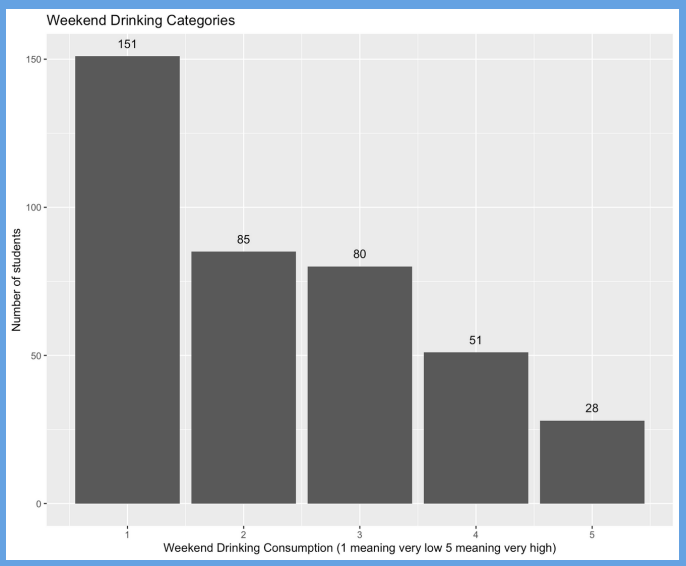


MATH STUDENTS

THROUGH THESE GRAPHS, IT IS EVIDENT THAT STUDENTS ON AVERAGE DRINK WHAT THEY CONSIDER "VERY LOW" ON **WEEKDAYS**, WHICH CORRELATES TO A SCORE OF 1. ALSO THERE IS VERY LITTLE DIFFERENCE BETWEEN THE SCORES OF PORTUGUESE STUDENTS AND MATH STUDENTS WHERE THE AVERAGE FOR PORTUGUESE STUDENTS BEING **1.50** AND MATH STUDENTS BEING **1.48**. THE GRAPHS SHOW THE DISTRIBUTIONS ALSO BEING VERY SIMILAR.



PORTUGUESE STUDENTS



MATH STUDENTS

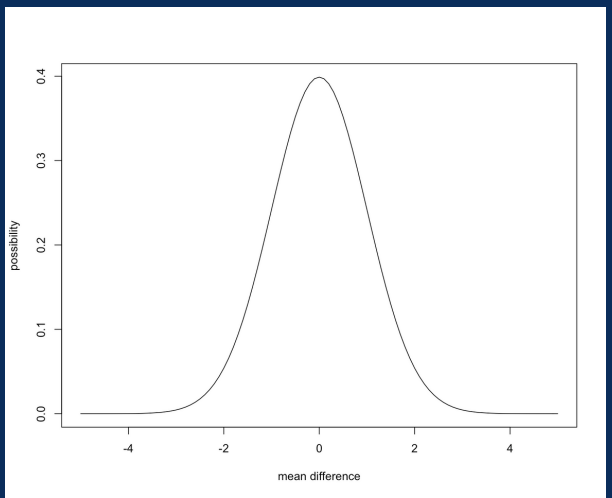
THROUGH THESE GRAPHS, IT IS EVIDENT THAT STUDENTS ON AVERAGE DRINK WHAT THEY CONSIDER "LOW" ON **WEEKENDS**, WHICH CORRELATES TO A SCORE OF 2. ONCE, AGAIN THERE IS VERY LITTLE DIFFERENCE BETWEEN THE SCORES OF PORTUGUESE STUDENTS AND MATH STUDENTS WHERE THE AVERAGE FOR PORTUGUESE STUDENTS BEING **2.28** AND MATH STUDENTS BEING **2.29**. THE GRAPHS SHOW THE DISTRIBUTIONS ALSO BEING VERY SIMILAR.

QUESTION: CAN WE SAY THAT STUDENTS DRINK MORE ON WEEKENDS THEN ON WEEKDAYS BASED ON THESE OBSERVATIONS?

HYPOTHESIS TESTING

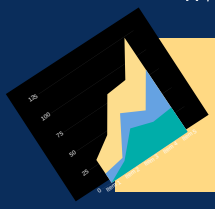
NULL HYPOTHESIS: PORTUGUESE STUDENT DRINKING ON WEEKENDS AND PORTUGUESE STUDENT DRINKING ON WEEKDAYS HAVE THE SAME DISTRIBUTION.

- MEAN FOR WEEKDAY: 1.50
- MEAN FOR WEEKEND: 2.28
- DIFFERENCE IN MEANS IS .78
- Z-SCORE: 12.524
- P: 0
  - THIS MEANS THAT THERE IS A 0% CHANCE THAT THE DIFFERENCE IN DRINKING SCORE BETWEEN WEEKDAY DRINKING AND WEEKEND DRINKING WOULD BE.78 OR HIGHER.
  - WE CAN REJECT THE NULL HYPOTHESIS SINCE P-VALUE IS LESS THAN 5%

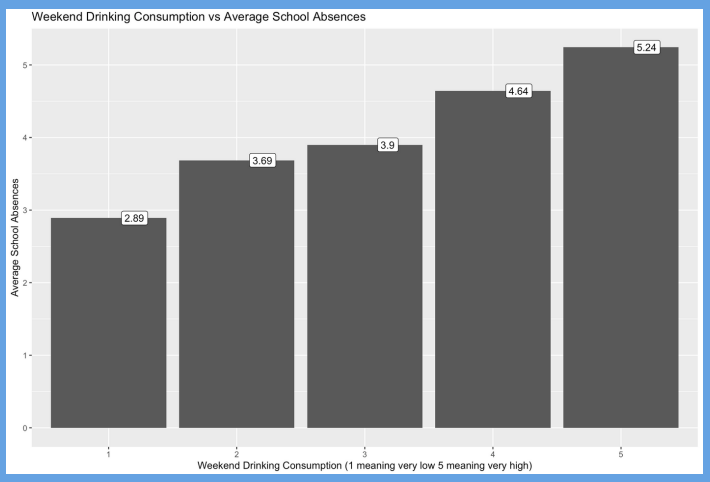


**ONLY USED PORTUGUESE STUDYING STUDENT DATA SINCE DISTRIBUTIONS WERE SIMILAR.

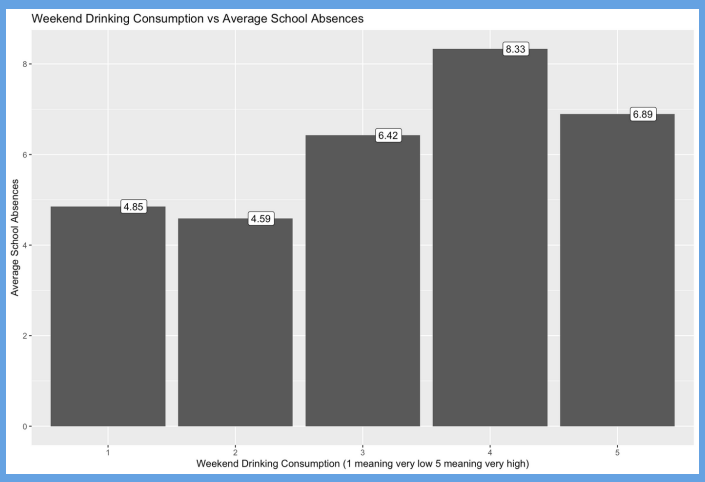
NOW, LET'S EXPLORE ANOTHER TOPIC: DOES THIS DATA SHOW THAT DRINKING CAN IMPACT ACADEMICS AS MANY SCHOOLS CLAIM?



MORE GRAPHS



PORTUGUESE STUDENTS



MATH STUDENTS

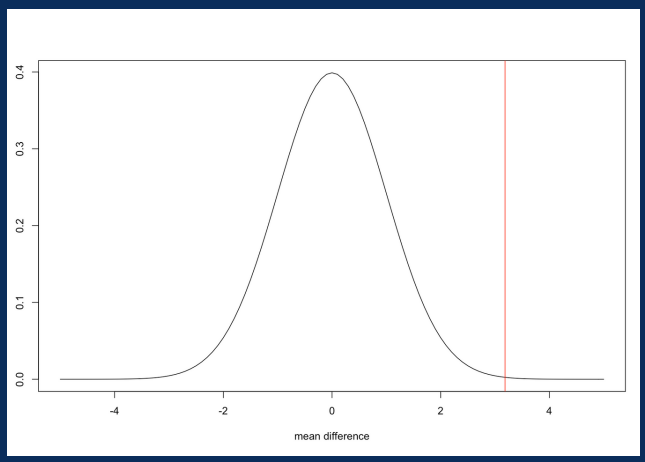
WE CAN SEE AN UPWARD TRNED BETWEEN DRINKING LEVEL ON WEEKENDS AND ABSENCES. SURPRISINGLY, THERE WAS A LARGE DIFFERENCE IN AVERAGE NUMBER OF ABSENCES FOR STUDENTS STUDYING MATH VS. PORTUGUESE. THE AVERAGE FOR MATH STUDENTS WAS AN AVERAGE NUMBER OF ABSENCES FOR MATH STUDENTS WAS **5.71** VS. **3.66** FOR PORTUGUESE STUDENTS. USING HYPOTHESIS TESTING (FOUND IN R FILE), WE FOUND THA THERE WAS A P VALUE OF **.000718%** AND THEREFORE WE CAN REJECT THE NULL HYPOTHESIS THAT BOTH GROUPS HAVE THE SAME DISTRIBUTION.

QUESTION: CAN WE SAY THAT STUDENTS WHO DRINK MORE HAVE MORE ABSENCES?
QUESTION: CAN WE SAY THAT STUDENTS WHO DRINK MORE HAVE WORSE GRADES?

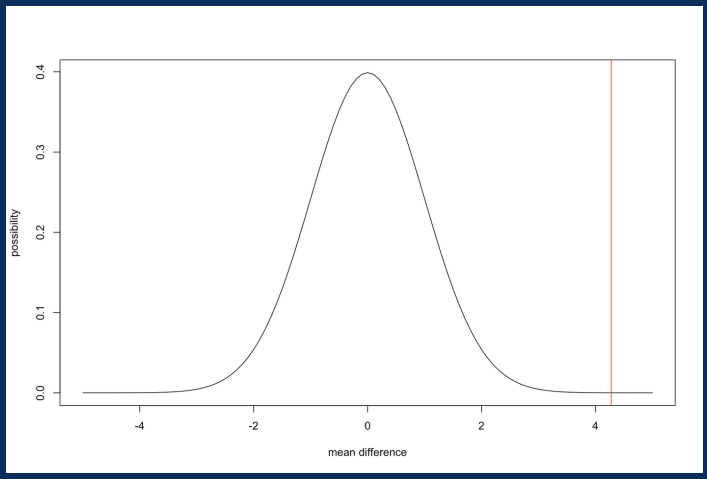
MORE HYPOTHESIS TESTING

NULL HYPOTHESIS: PORTUGUESE STUDENTS WITH A WEEKEND DRINKING LEVEL <= 3 AND PORTUGUESE STUDENTS WITH A WEEKEND DRINKING LEVEL > 3 HAVE THE SAME DISTRIBUTION FOR ABSENCES.

- MEAN ABSENCES FOR > 3: 4.85
- MEAN ABSENCES FOR <=3: 3.35
- DIFFERENCE IN MEANS IS 1.5
- Z-SCORE: 3.18
- P: .073%
  - THIS MEANS THAT THERE IS A .073% CHANCE THAT THE DIFFERENCE IN ABSENCES FOR THE TWO GROUPS WOULD BE 1.5 DAYS OR HIGHER.
  - WE CAN REJECT THE NULL HYPOTHESIS SINCE P-VALUE IS LESS THAN 5%



NULL HYPOTHESIS: PORTUGUESE STUDENTS WITH A WEEKEND DRINKING LEVEL <= 3 AND PORTUGUESE STUDENTS WITH A WEEKEND DRINKING LEVEL > 3 HAVE THE SAME DISTRIBUTION FOR GRADES (SCALE OF 1-20 FOR GRADES).



- MEAN GRADE FOR > 3: 12.17
- MEAN GRADE FOR <=3: 10.87
- DIFFERENCE IN MEANS IS 1.3
- Z-SCORE: 4.278
- P: .00095%
  - THIS MEANS THAT THERE IS A .00095% CHANCE THAT THE DIFFERENCE IN GRADES FOR THE TWO GROUPS WOULD BE 1.2 POINTS OR HIGHER.
  - WE CAN REJECT THE NULL HYPOTHESIS SINCE P-VALUE IS LESS THAN 5%

KEEP IN MIND: ALTHOUGH HYPOTHESIS TESTS DO NOT ANSWER OUR QUESTIONS DIRECTLY, THEY DO TELL US THAT DISTRIBUTIONS BETWEEN LOW LEVEL DRINKERS AND HIGH LEVEL DRINKERS ARE DIFFERENT

But Which Students Are More likely to Be Heavy Drinkers ?

1) STUDENTS WHO ONLY STUDY 1 HOUR A WEEK ARE MORE LIKELY TO BE HEAVY DRINKERS:
   - PRIOR PROBABILITY: PR(HEAVY DRINKERS) = **20.34%**
   - POSTERIOR PROBABILITY: PR(HEAVY DRINKERS | STUDIES 1 HR A WEEK) = **32.55%**
   - SUPPORT: PR(HEAVRY DRINKER & STUDIES 1 HR A WEEK) = **69**

2) STUDENTS WHO GO OUT AT A HIGH RATE WITH FRIENDS ARE MORE LIKELY TO BE HEAVY DRINKERS:
   - PRIOR PROBABILITY: PR(HEAVY DRINKERS) = **20.34%**
   - POSTERIOR PROBABILITY: PR(HEAVY DRINKERS | GO OUT LEVEL > 3) = **37.85%**
   - SUPPORT: PR(HEAVRY DRINKER & GO OUT LEVEL >3) = **95**

3) STUDENTS WHO FAIL AT LEAST 1 CLASS AND GREW UP IN AN URBAN AREA ARE MORE LIKELY TO BE HEAVY DRINKERS:
   - PRIOR PROBABILITY: PR(HEAVY DRINKERS) = **20.34%**
   - POSTERIOR PROBABILITY: PR(HEAVY DRINKERS | ((FAILED CLASS > 1) & (GREW UP IN UBRAN AREA)) = **33.33%**
   - SUPPORT: PR(HEAVRY DRINKER & GO OUT LEVEL >3) = **20** (IN CLASS, WE NEEDED >=15)

Is The Data Reliable?

- THE TWO DATASETS BOTH HAVE 33 VARIABLES AS WELL AS 649 ROWS OF DATA FOR THE PORTUGUESE STUDYING STUDENTS AND 395 ROWS FOR THE MATH STUDYING STUDENTS. THIS MEANS THAT THE AMOUNT OF DATA IS SUFFICIENT.
- UCI MACHINE LEARNING IS A WELL REGARDED SOURCE OF DATA SO THE DATA CAN BE CONSIDERED RELIABLE.
- HOWEVER, HERE ARE SOME SIGNIFICANT SETBACKS IN HOW THE DATA WAS COLLECTED:
  - A LOT OF THE METRICS HAD A SCALE OF 1-5 INSTEAD OF SPECIFIC QUANTITIES. THIS BECOMES VERY VAGUE. FOR INSTANCE A 5 CORRELATING TO HEAVY DRINKING IS VAGUE SINCE PEOPLE HAVE DIFFERENT DEFINITIONS OF "HEAVY DRINKING"
  - THE DATA SHOULD HAVE BEEN INCLUSIVE OF ALL MAJORS RATHER THAN JUST MATH AND PORTUGUESE STUDYING STUDENTS TO MAKE SURE THAT ALL BACKGROUNDS WERE BEING REPRESENTED.
  - BEING IN SCHOOL COULD STILL MEAN THAT MANY ARE HESITANT TO SAY THAT THEY ARE HEAVY DRINKERS.

CONCLUSION

- STUDENTS AT THIS SCHOOL IN PORTUGAL CONSIDER THEMSELVES LOW AND VERY LOW LEVERL DRINKERS.
- DISTRIBUTIONS BETWEEN HEAVY DRINKING AND ABSENCES AND GRADES ARE NOT SIMILAR.
- STUDENTS WHO STUDY LITTLE, GO OUT A LOT WITH FRIEDS, AND FAIL 1 CLASS & GREW UP IN URBAN AREAS ARE LIKELY TO DRINK MORE.

ALTHOUGH THIS DATA CANNOT BE DIRECTLY APPLIED TO OTHER HIGH SCHOOLS AND COLLEGES, IT HELPS US REALIZE THAT THIS IS THE TYPE OF DATA THAT CAN HELP MAKE MORE IN DEPTH CLAIMS ABOUT DRINKING PATTERNS.

WITH DEEPER FINDINGS AND VISUALIZATIONS TO BACK UP CLAIMS SUCH AS "1 OUT OF 5 STUDENTS DO NOT DRINK", STUDENTS WOULD GAIN TRUST BEHIND WHAT SCHOOLS ARE SAYING AND PERHAPS EVEN FORM DIFFERENT OPINIONS ABOUT THE REALITY OF THE DRINKING SCENE IN COLLEGE.