

BIOS 664 Group F Final Project Presentation

# Availability of Study Rooms in UNC Libraries

Group F: Morgan Watson, Sophie Salem,  
Maya Yin, Swaycha Goli, Maya  
Krishnamoorthy

# Estimation Problem and Summary

- Goal: Estimate the proportion of available rooms in UNC libraries
- We chose to analyze the three libraries that have the most reservable rooms:  
Davis Library, the Undergraduate Library, and Health Sciences Library



# Our Population



## What population is to be sampled?

- Reservable rooms in Davis, the UL, and HSL

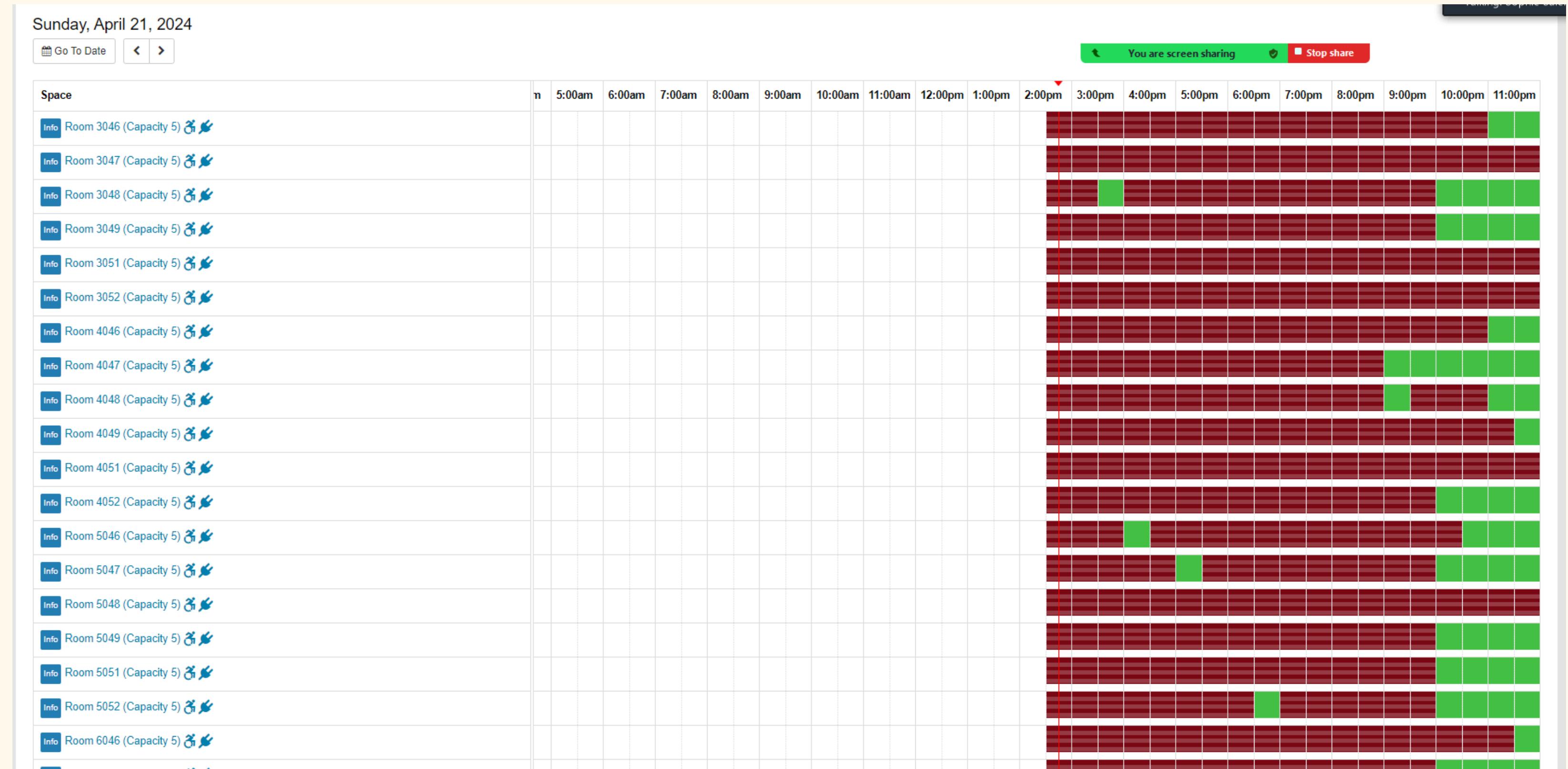
## What are the members of that population?

- Reservable rooms within a designated 1-hour time block between 8:00AM-8:00PM

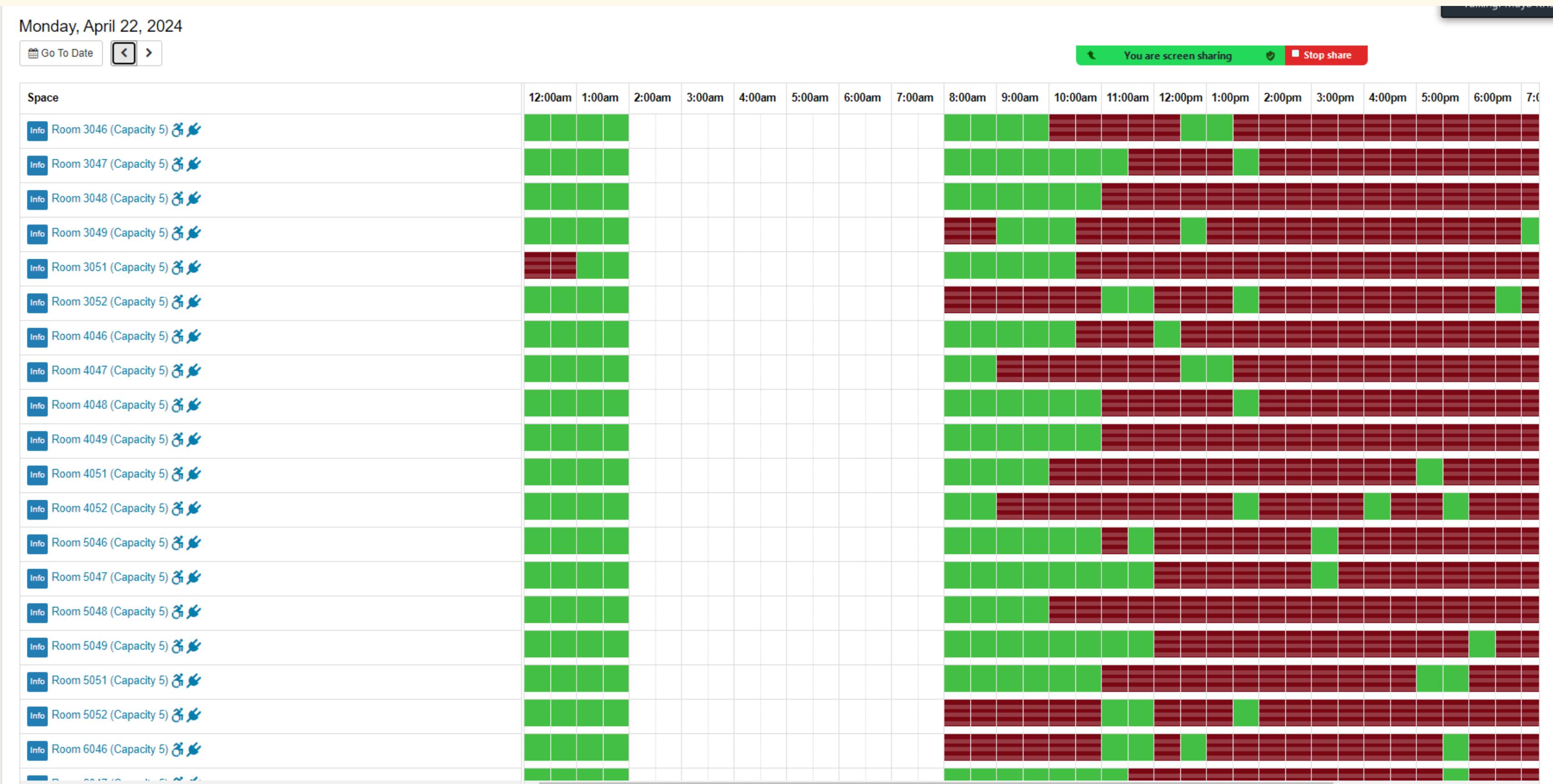
## What units are to be observed/measured, and what observations/measurements are to be made?

- We will measure members of the population and note whether the room is reserved or not, 24 hours in advance

# Measuring on the Day



# Measuring 24 hrs Before



# Sample Design Summary



# Sampling Frame

**How will you derive your sampling frame for each stage of selection?**

- Created an Excel sheet with all reservable rooms in each of the three libraries for each hour time block between 8:00AM and 8:00PM



**What will be the sampling unit(s), and how will they be selected?**

- One sampling unit is a reservable room at a specific hour time block
- Selected through the surveyselect procedure in SAS
- Sample was stratified by both library and time of day utilizing a proportional allocation design

**Will selection probabilities be the same for each member of the sample?**

- Due to the EPSEM design of our sample achieved through stratification with a proportional allocation the selection probabilities should be the same
- Selection probabilities for the final sample ranged from 0.077-0.079 likely as a result of rounding sample sizes to a whole number considered approximately EPSEM

# Design Summary Table



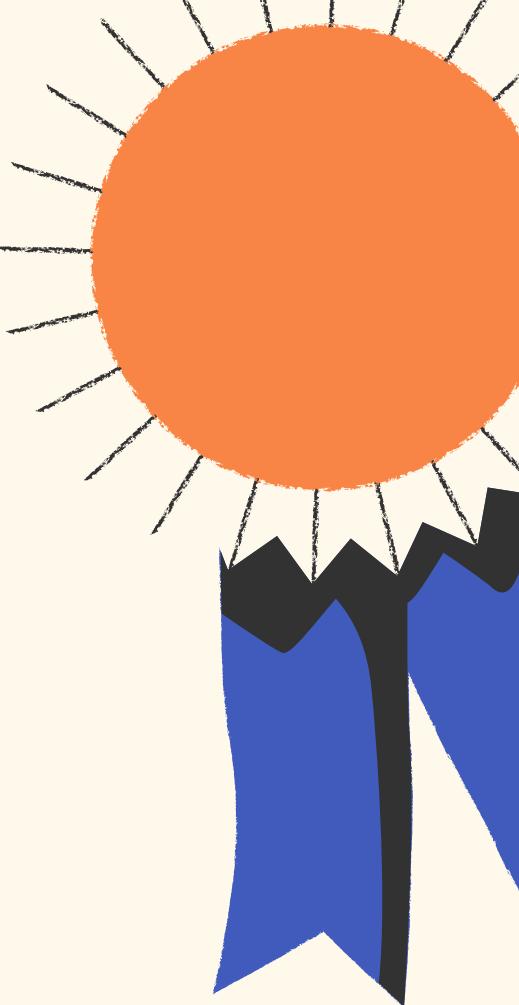
Stage	Description of Sampling Unit (What is being sampled?)	Stratification (By what? How many strata? Is sample allocation balanced, proportionate and/or disproportionate?)	Type of Selection Method Used to Choose Sampling Units (e.g. SRS, PPSWOR, etc.)	Sample Size (How many sampling units were chosen?)
1	A study room during a one-hour interval on a weekday	9 strata formed by crossing 3 libraries by 3 time periods. The sample is allocated proportionally across strata.  Strata: HSL in the morning, mid-day, and evening; the same applies for Davis and the UL	SRS	357

# Sample Size Summary

- Controlling margin of error at 0.05 using a z value of 1.96
- Determine the number of samples allocated to each stratum proportional to the stratum size



# Sample Size Justification



## Population size:

- 77 study rooms \* 12 one-hour intervals \* 5 weekdays = 4620 units

## Variance estimation:

- a conservative estimation  $p(1-p)=(0.5)(0.5)=0.25$

## Sample size calculation specifying $e=0.05$ and $z=1.96$ :

$$n' = \frac{z_{1-\frac{\alpha}{2}}^2 s^2}{e^2} = \frac{1.96^2 (0.25)}{0.05^2} = 384.16$$

Assume SRS with replacement

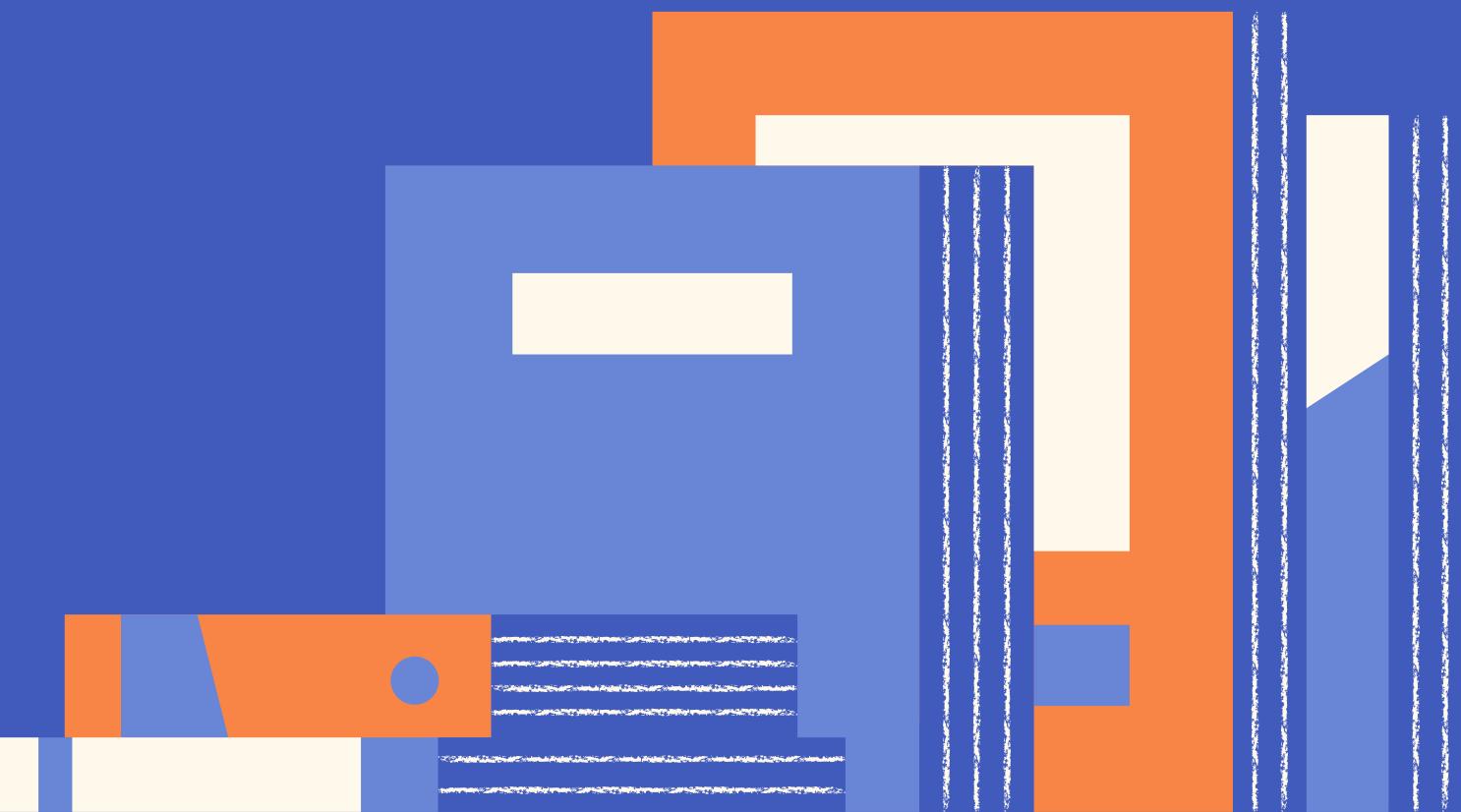
$$n = \frac{n'}{1 + \frac{n'}{N}} = 354.67$$

Adjust for finite population correction

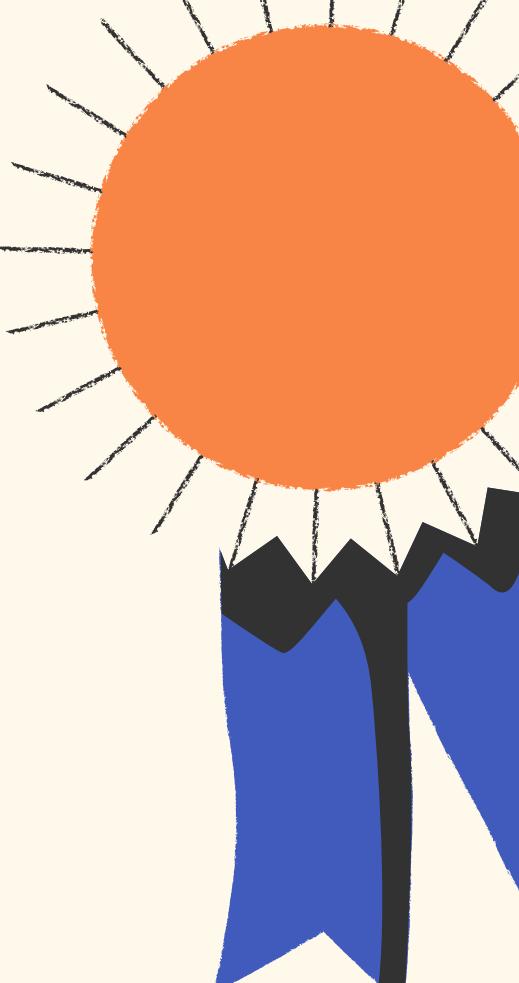
Library	Davis library			Undergraduate Library			Health Science Library			Total
Time interval										
Stratum Size (Nh)	860	860	860	240	240	240	440	440	440	4620
Sample Size (nh)	66	66	66	19	19	19	34	34	34	357

- Stratum size = rooms in a library \* 4 hours interval \* 5 days
- Strata sample size = total sample size \* (stratum size)/(population size)
- Strata sample sizes were rounded up to ensure ample sampling
- DEFF is estimated to be less than or equal to 1 due to no clustering

# Data Collection and Data Processing Summary



# Sample Selection Procedure



## 1. Proportionally allocated sample size for each library

- Using calculations from our sample size formula
- Davis - 66, UL - 19, HSL - 34 per time interval per day

## 2. Stratify on time interval

- Using SAS
- Proportional allocation

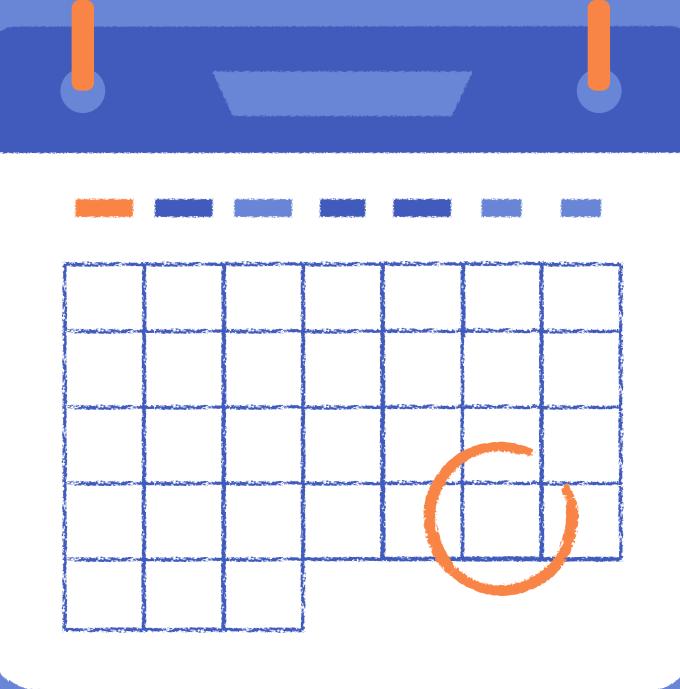
## 3. Select sampling unit via SRS

- Sample chosen from each time interval was proportional to stratum size

```
/*Step 1: sort HSL data by stratification variable (time interval)*/
proc sort data=hslsrs; by Time_Interval; run;

/*Step 2: Select PROPORTIONATE stratified sample for HSL (stratified by 'time interval') */
proc surveyselect data=hslsrs method=srs n=34 out = hsl_sample seed = 012620241 stats;
  strata Time_Interval;
run;
```

# Sample Measurements



## Sampling Frame

- Entered each study room in each library into an Excel sheet
- 60 repetitions for each room: 12 time slots x 5 days

## Pilot Testing

- All group members practiced taking the sample at the same time to ensure fairness
- The sample of reservable study spaces was split into 12 hourly time slots between 8:00am and 8:00pm
- 24 hours before the time interval of interest, we logged into <https://calendar.lib.unc.edu/reserve/> to check the availability of the study space (available = 1 and unavailable = 0)

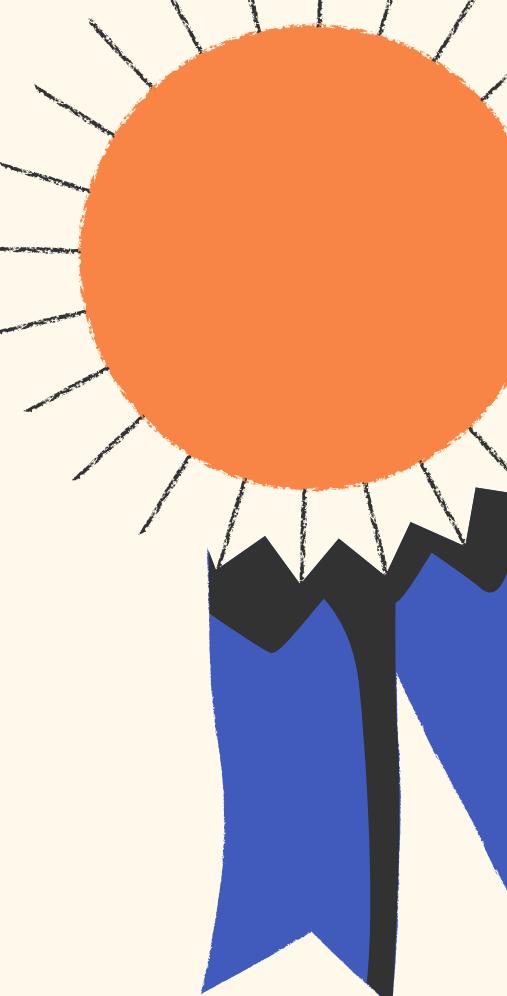
## Data Collection per Day

- We collected data between March 20th and 26th (20, 21, 24, 25, 26)
- Each group member checked the website and manually entered the availability into an Excel sheet

# Example Data Collection

1	Davis Library	Cube 4	1pm-2pm	Afternoon	1
1	Davis Library	5052	1pm-2pm	Afternoon	0
1	Davis Library	8051	1pm-2pm	Afternoon	0
1	Davis Library	7051	1pm-2pm	Afternoon	0
1	Davis Library	7052	1pm-2pm	Afternoon	0
1	Health Sciences Library	425	1pm-2pm	Afternoon	0
1	Health Sciences Library	428A	1pm-2pm	Afternoon	1
1	Davis Library	6048	2pm-3pm	Afternoon	0
1	Davis Library	8047	2pm-3pm	Afternoon	0
1	Davis Library	8051	2pm-3pm	Afternoon	0
1	Davis Library	7047	2pm-3pm	Afternoon	0
1	Davis Library	3051	2pm-3pm	Afternoon	0
1	Davis Library	4051	2pm-3pm	Afternoon	0
1	Undergraduate Library	220	2pm-3pm	Afternoon	0
1	Undergraduate Library	221	2pm-3pm	Afternoon	0
1	Davis Library	4046	3pm-4pm	Afternoon	0
1	Davis Library	5051	3pm-4pm	Afternoon	0
1	Undergraduate Library	202A	3pm-4pm	Afternoon	0
1	Health Sciences Library	428C	3pm-4pm	Afternoon	0

# Sources of Measurement Error



1

We anticipated the problem of rooms being mostly unavailable if we checked on the day of interest, so we had decided in our original design plan to check the availability of rooms 24 hours before.

2

On Fridays, all library reservations stop at 5:00 pm, so we had to mark them as missing data. We decided to analyze them as missing rather than unavailable because we did not anticipate that in our design plans.

3

We had spelling errors in the time interval variable for “Afternoon,” so we checked them before completing the weighting and analysis tasks.

4

Not every sampling unit was checked for availability exactly 24 hours before.

# Weighting Summary

- More concerned about non-response bias than coverage bias
- ASSUMPTION: non-respondents have the same distribution as respondents
- Weighting class method
- Weight class = library \* time interval (strata)
  - Assigned base weights
  - Found total weights in each weight class
  - Found proportion of respondents in each weight class
  - Calculated psi values (weight of respondents/total weight)
  - Calculated final non-response weights

**Weights (no missing data)**

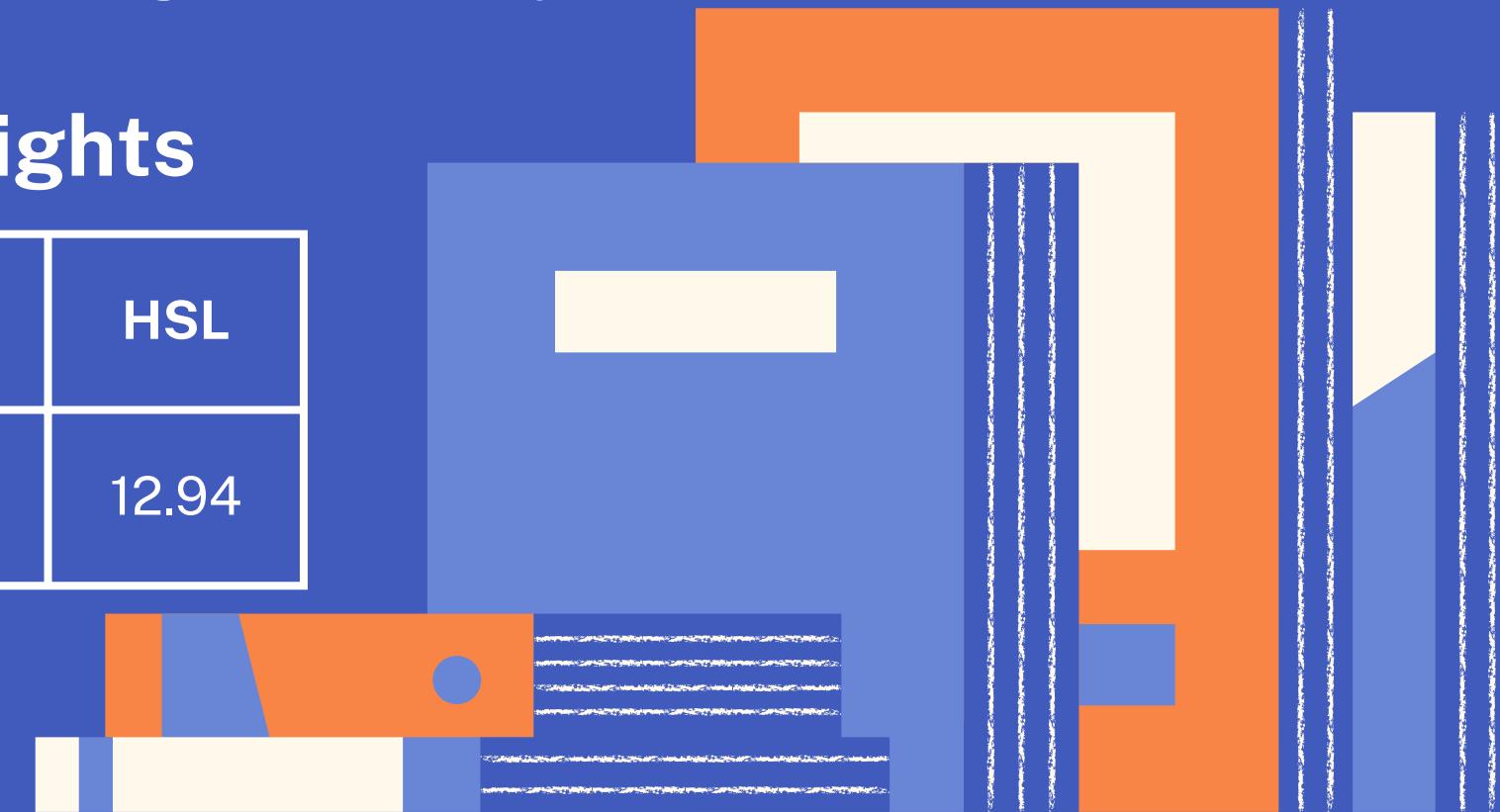
	Morning	Afternoon	Evening	Totals
Davis	17.71 18.62	20.51 18.62	18.34 18.62	56.56 55.85
UL	4.82 5.19	5.72 5.19	5.12 5.19	15.66 15.58
HSL	8.02 9.52	10.49 9.52	9.26 9.52	27.78 28.57

\*top number = weight before NR adjustment

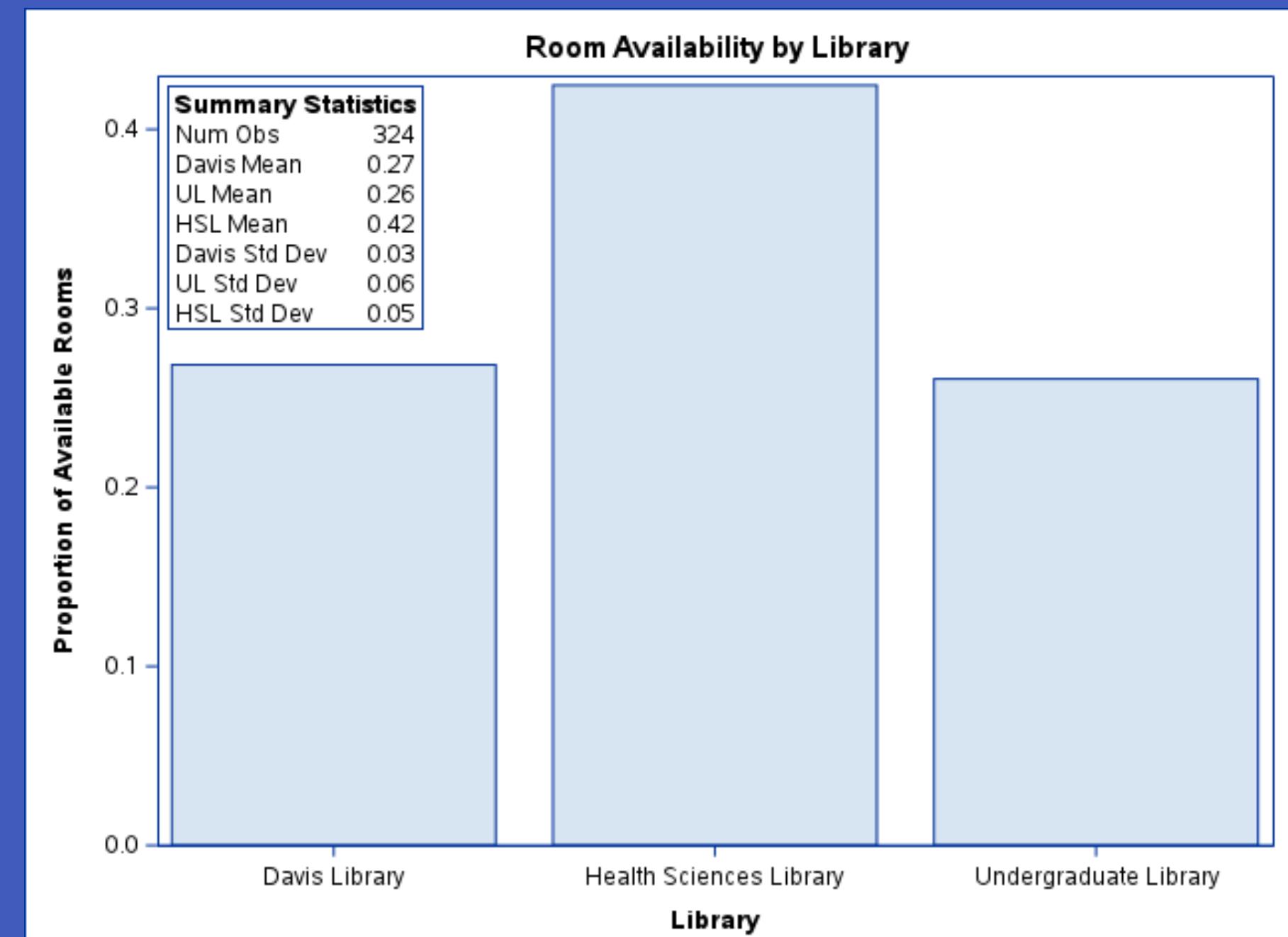
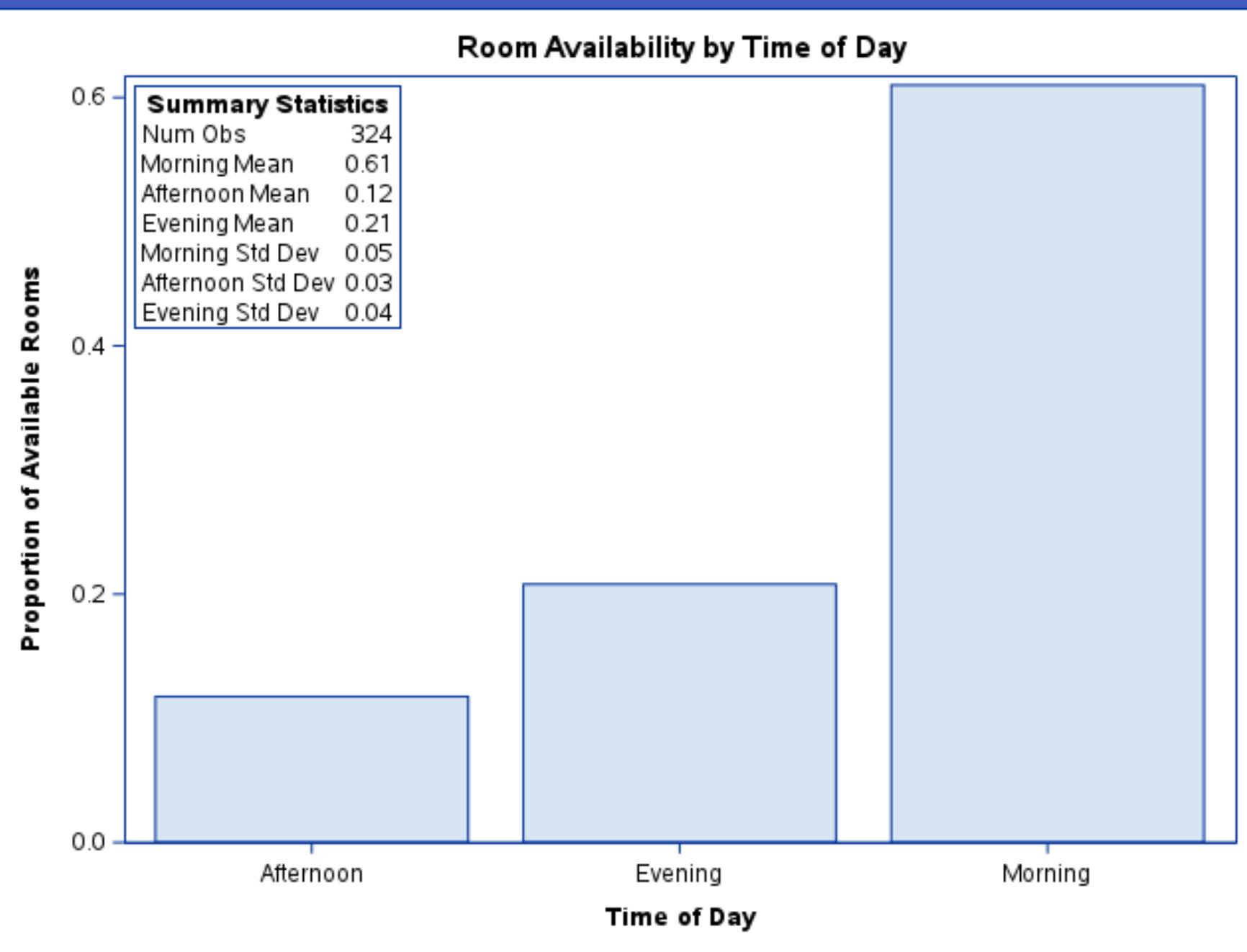
\*bottom number = weight after NR adjustment

## Base Weights

Davis	UL	HSL
13.03	12.63	12.94

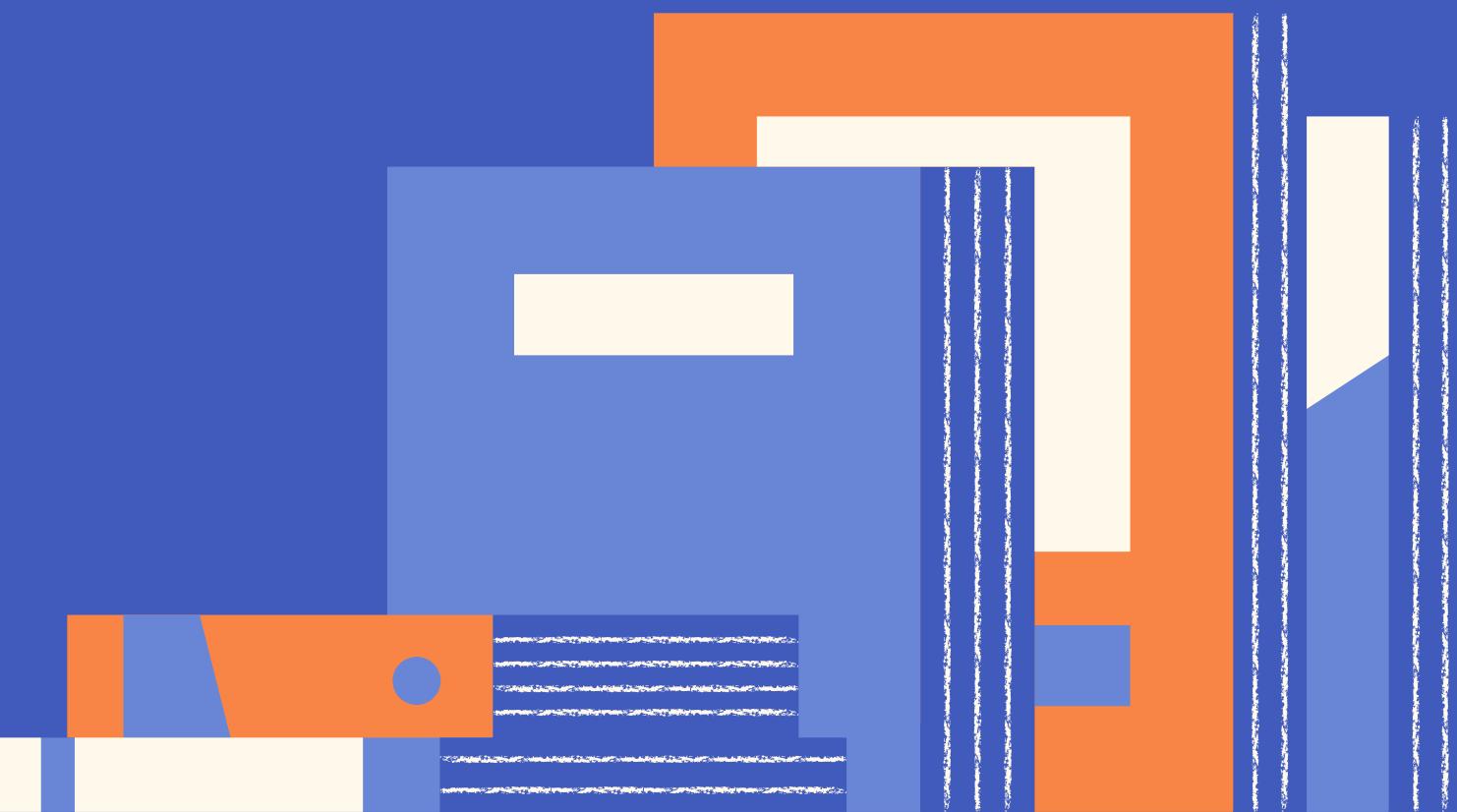


# Analysis Summary



# Final Conclusions and Takeaways

- Overall, 31% of study rooms were available across the three libraries and time periods
- Health Sciences Library had the most availability
- Time period 8 am - 12 pm had the most available rooms



# Error and

## Bias

## Analysis



### Non-sampling errors:

- Frame problem missing Friday evening data
- Measurement errors:
  - Failure to refresh the reservation page during data collection
  - Data collected not exactly 24 hours before

### Accuracy and precision of measurement:

- MOE successfully controlled within 0.05
- The majority of missing data are from HSL in the morning

# Thank you!

Questions?