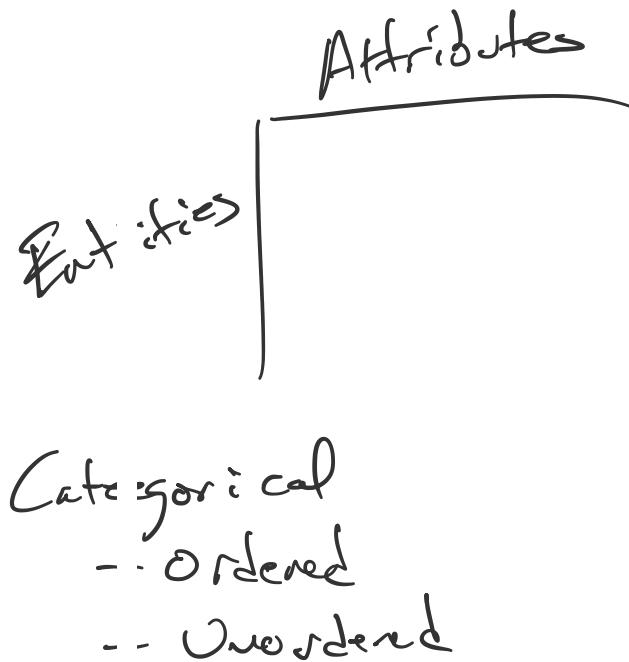


Lesson Plan 2/1

Thursday, February 1, 2018 9:23 AM

- Admin
 - o Class notes linked to in Calendar
 - o Any issues, comments?
- Hector's Data Science Corner
 - o Talk at the HCIL: https://hcil.umd.edu/events/event/bbl-bahador-saket-georgia-t_ech/
 - o Data Skeptic: <https://dataskeptic.com/>
- Something I missed from last time
 - o Categorical attributes and 'factors'
- More operations
 - o Sorting: arrange
 - o Mutate
 - o Summarize
 - o Group by
- A couple other R programming issues
 - o Vectors
 - o Attributes are vectors
 - o Defining functions
- Plotting
 - o The ggplot triumvirate:
 - Data
 - Mapping
 - Representation
 - o Mappings
 - o Representations
 - o Frequently Used Plots
 - Scatter plot
 - Bar graph
 - Histogram
 - Boxplot
- Rmarkdown
 - o Markdown
 - o Code chunks



- Code Editors
- Notebook
- Metadata
- Rendering

Numerical

-- Discrete

-- Continuous

- # of students "Yes" = Discrete

- Avg # of students: "Yes": Continuous

Day

M	5
T	10
W	12
Th	13

avg # Yes

site

Samp	6.3
HSP	7.2
CSR	100

"factor" in R: Categorical Data

Turns a string vector →

vector of categorical values

possible values of a factor, are called "levels"

Operations:

Attributes: select

Entities: slice, filter, sample_x

Pipe : '%>%'

Ordering: Arrange entities by attribute value

Summary:

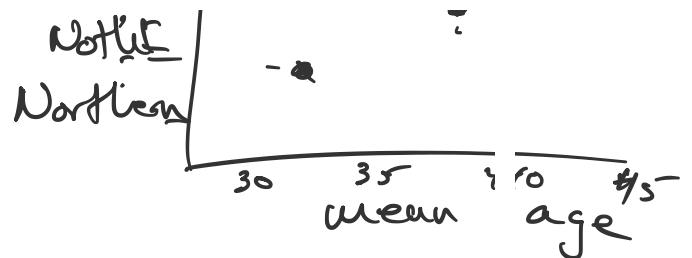
grouping:

Create new : 'mutate'
attributes



ggplot : grammar of graphics





- 1) Data: Entities & attributes
- 2) Mapping: Data values \rightarrow graphical characteristics
- 3) Representation: Geometrically representing characteristics

What is 'aes', aesthetic characteristics