|  |  |
| --- | --- |
|  |  |
| Opening Slide | Good morning.  Welcome to Day 2 of Computational Text Analysis Week.  My apologies, my voice is weak, so I am trying to save it for answering your questions.  Anyways, for any of you who were not with us yesterday, my name is Jeremy Mikecz  And I work with the Research Data Services team here at the library. My colleagues Lora Leligdon and Simon Stone are also here to help.  Fortunately, most of our materials are designed for self-guided instruction.  So, after this opening presentation, we will move to some hands-on practice. |
| Text data | We often associate the concept of data with numerical and quantitative data.  However, data are really any type of information that can be read and interpreted by a computer. With modern technology, we can include text, images, audio, and video all as data as well as numerical data tables.  Text analysis is thus an important field because the majority of data being produced is texts created by humans for humans…  … ahem… well until this past year and the development of AI text generation tools like ChatGPT |
| CTA | What we are calling Computational text analysis goes by many names…  In literary studies it is called distant reading (in contrast to close reading)  In the humanities: it is considered part of Cultural Analytics (in contrast to data analytics)  In computer science: natural language processing (vs. programming languages)  And you can see some other examples here |
| Types of texts | Computational text analysis can mean many different things…  Depending on the type of text for example: has it been cleaned of mistakes, have the spellings and grammar of historical texts been standardized?  Is it a largely unstructured text like a novel or journal diary? Or is it structured or formulaic like a report or encyclopedia entry? |
| Types of analysis | There are many different |
|  |  |