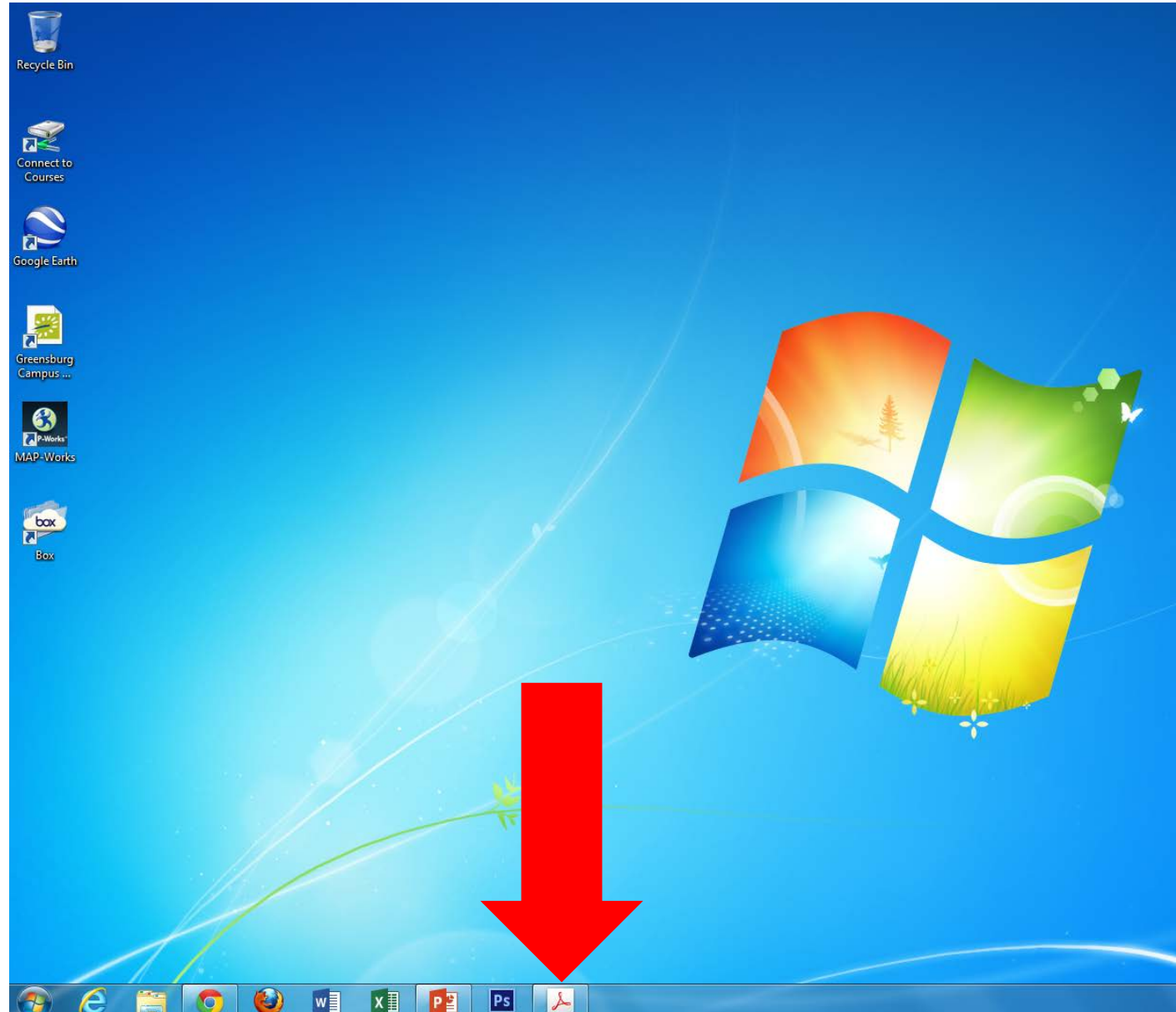


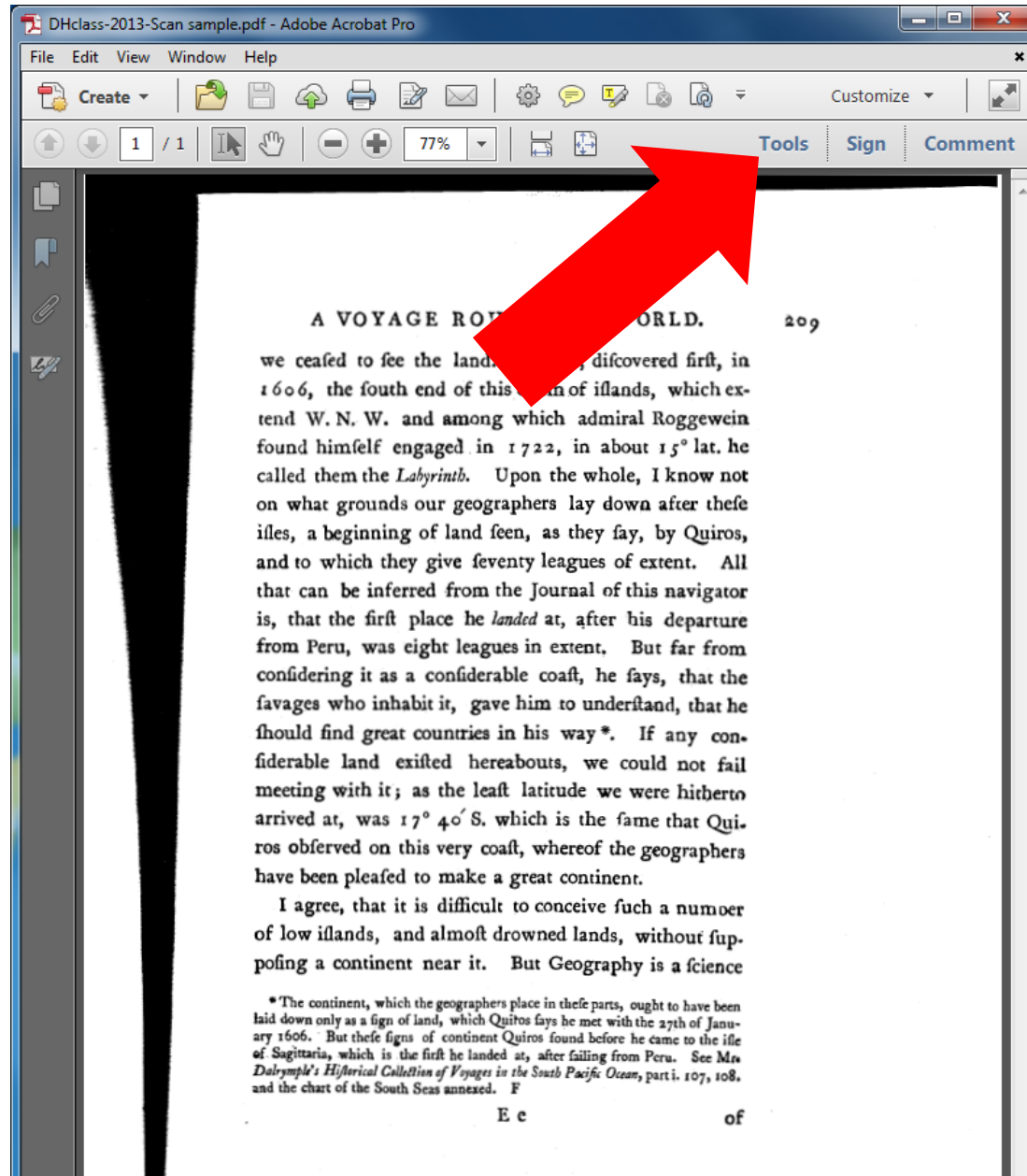
# DH/DS OCR Homework Assignment

- Due Monday Sept 9

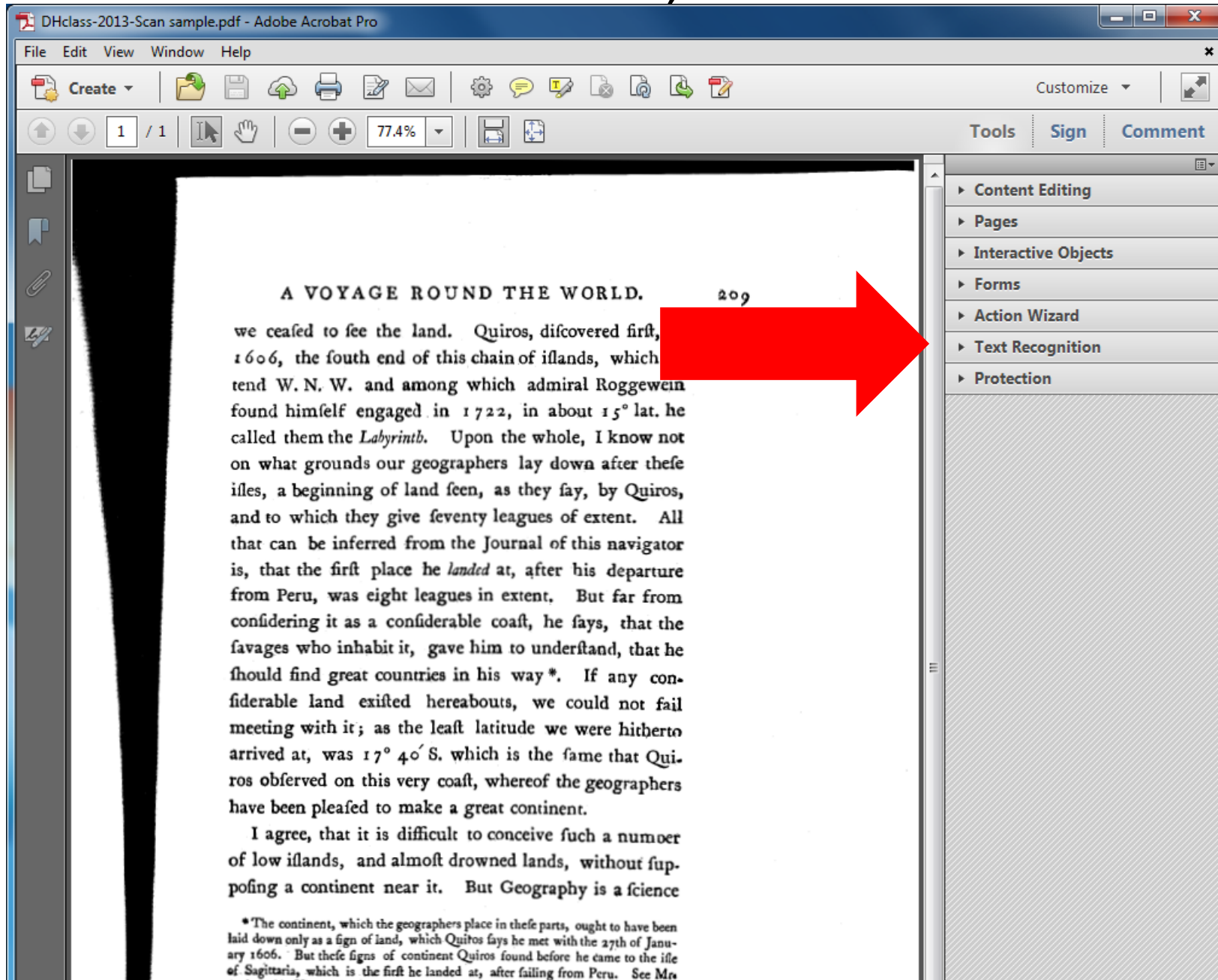
1) Download and Open both of the PDF files on the Syllabus web page (or in Box) using Adobe **Acrobat Pro** (*not* Reader)



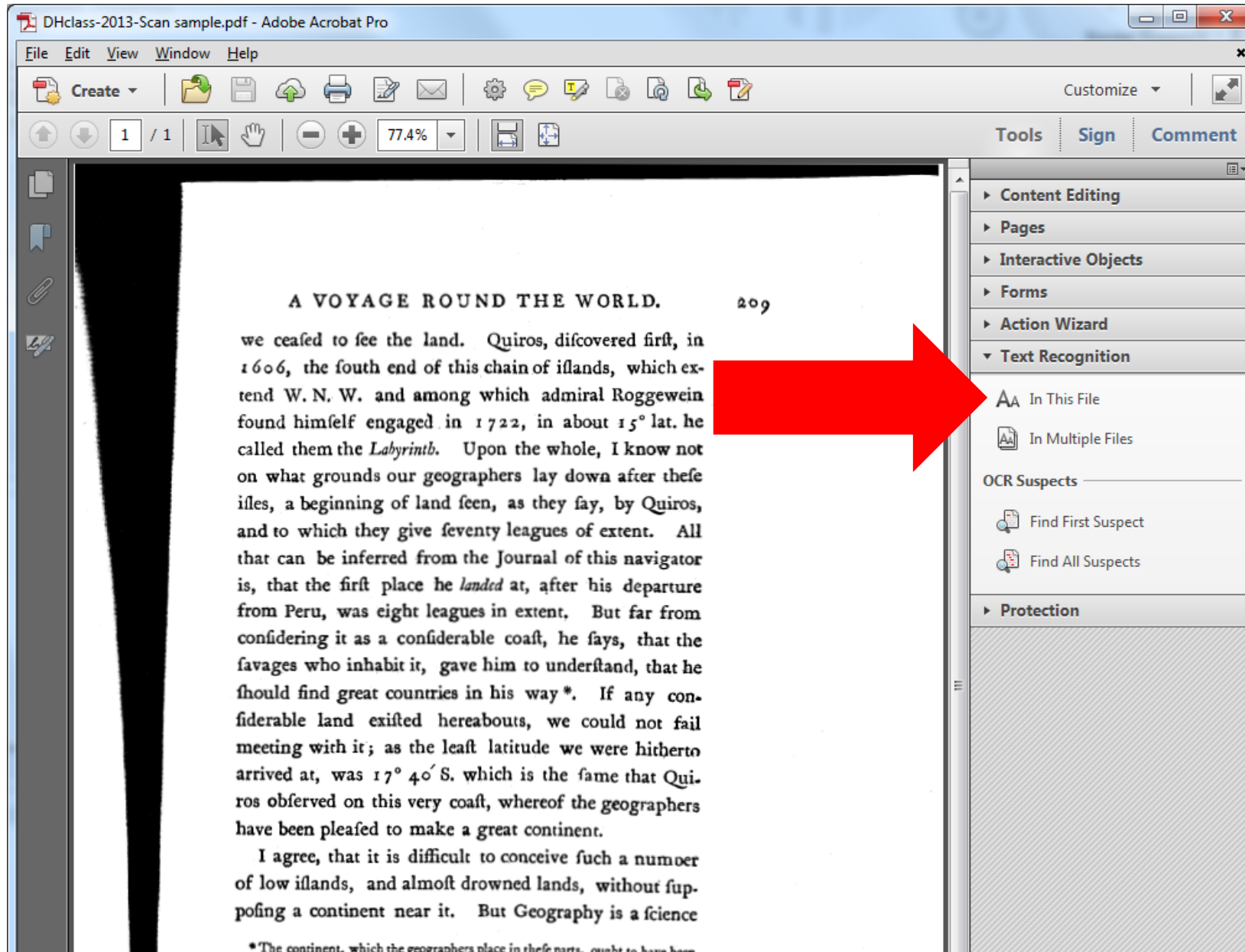
## 2) In Adobe Acrobat Pro, click on the Tools menu



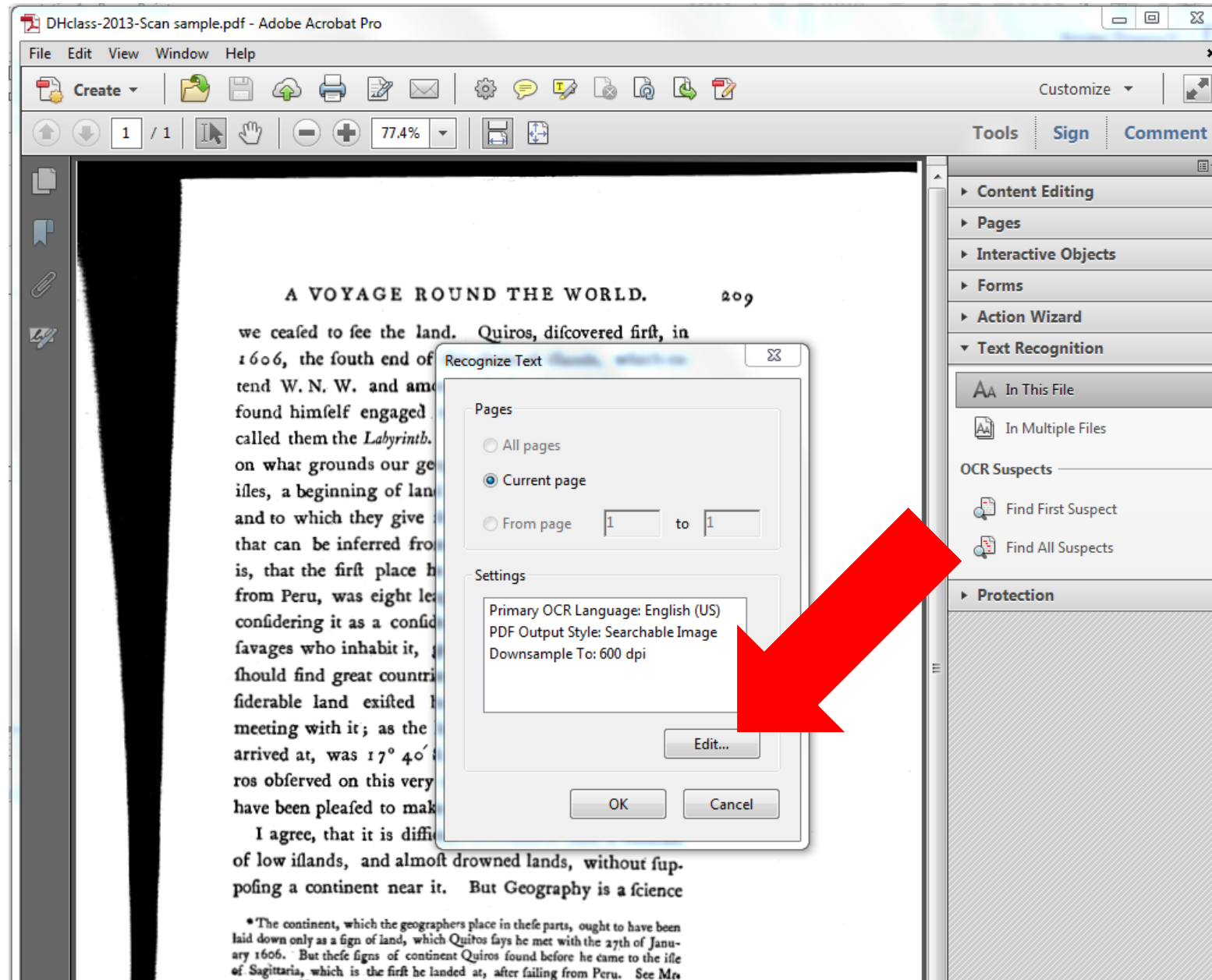
### 3) Under the Tools menu, select Text Recognition



## 4) Under Text Recognition, select In This File

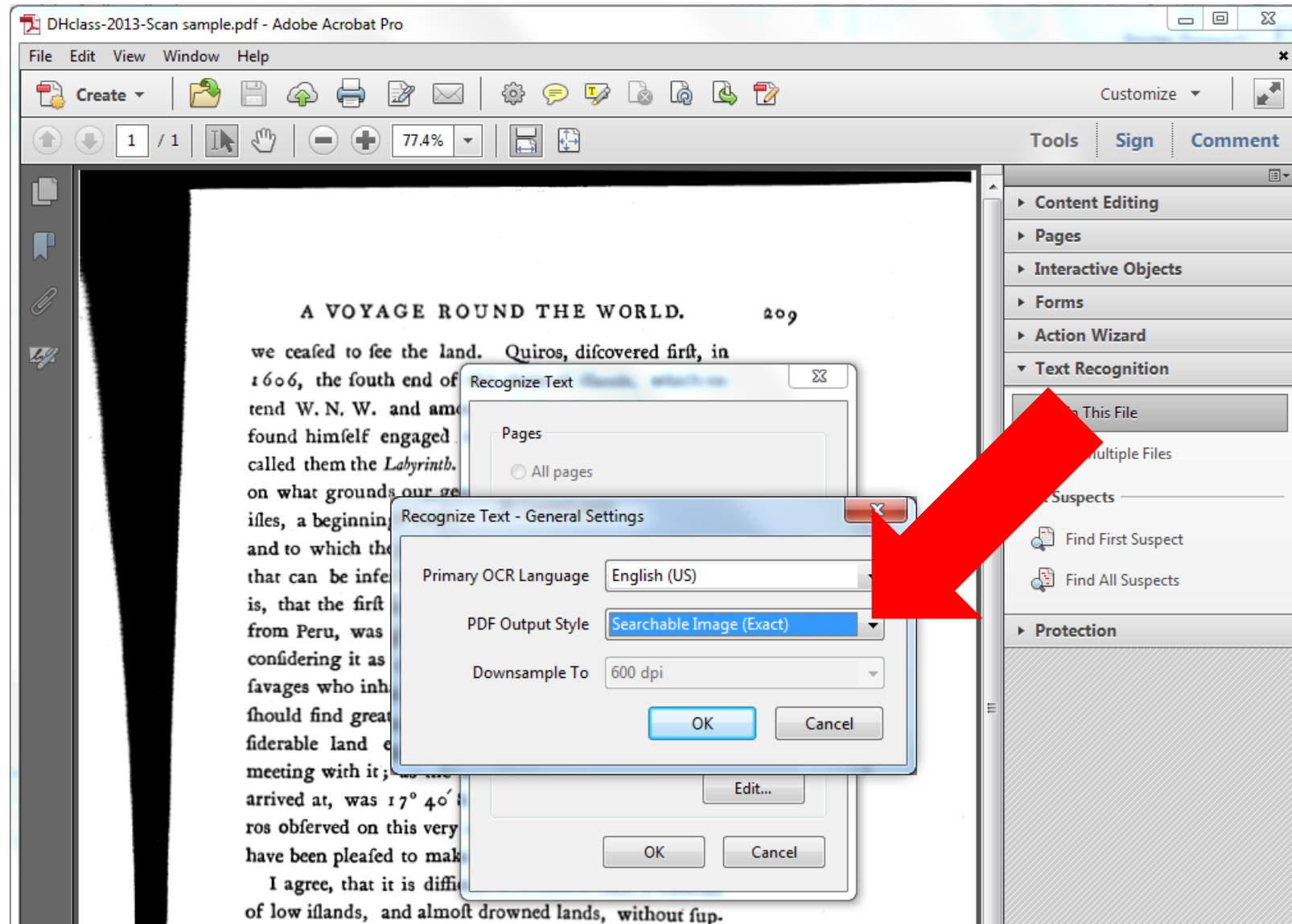


## 5) In the Recognize Text dialog box, click on Edit

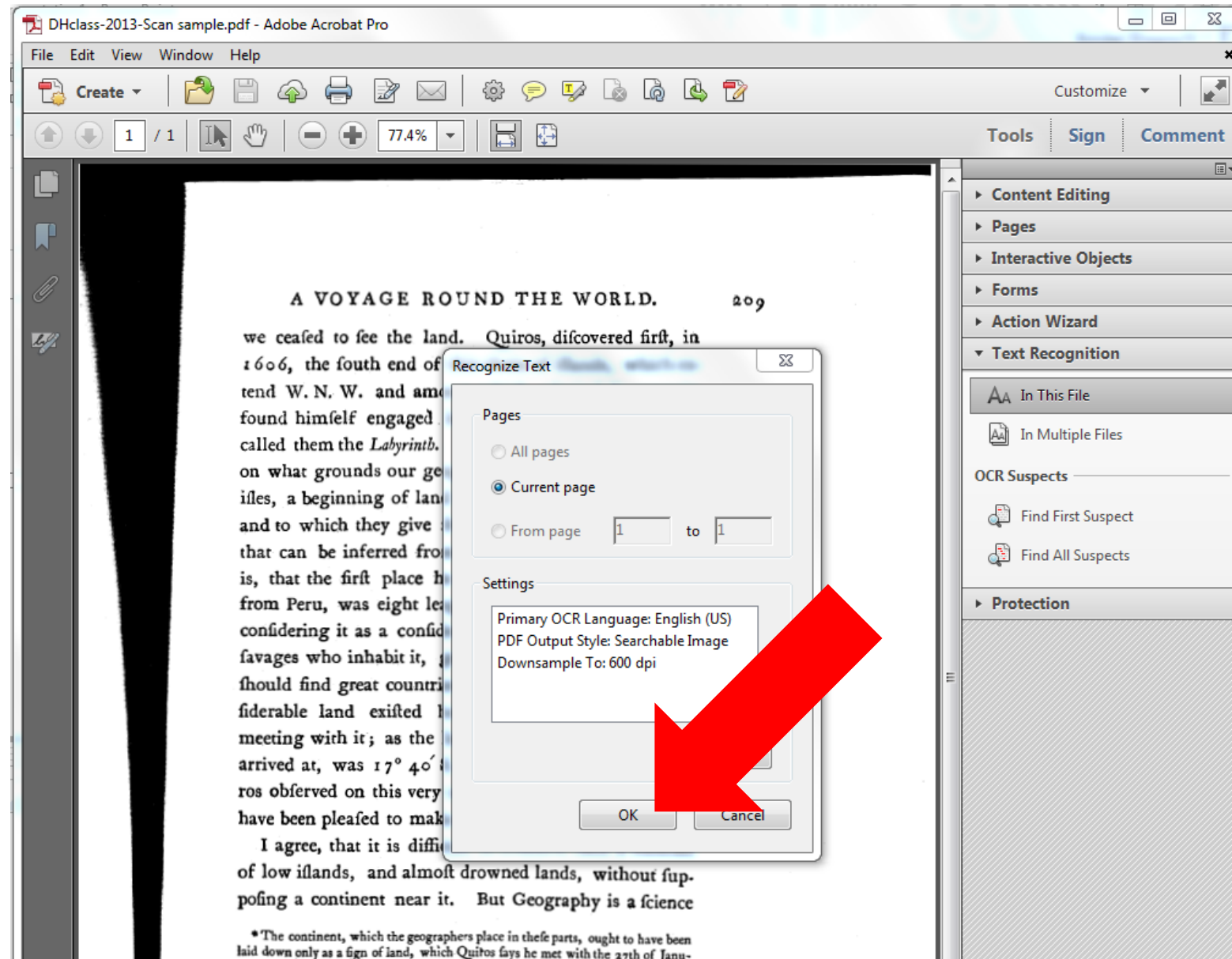




- 6) In the Recognize Text – General Settings dialog box:
- Change PDF Output Style to Searchable Image (Exact)
  - Click OK

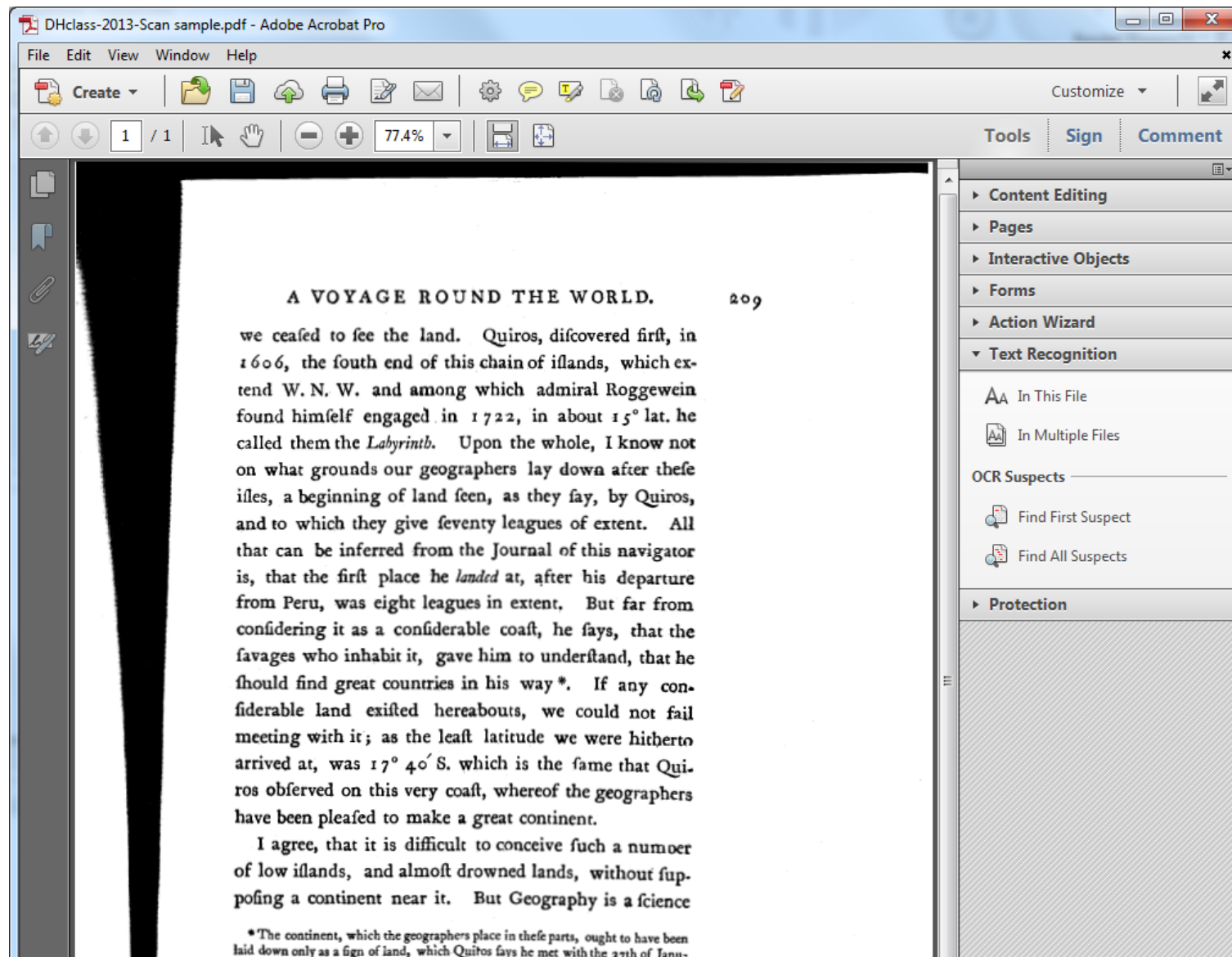


7) Click OK on the Recognize Text dialog box to perform OCR on this document.

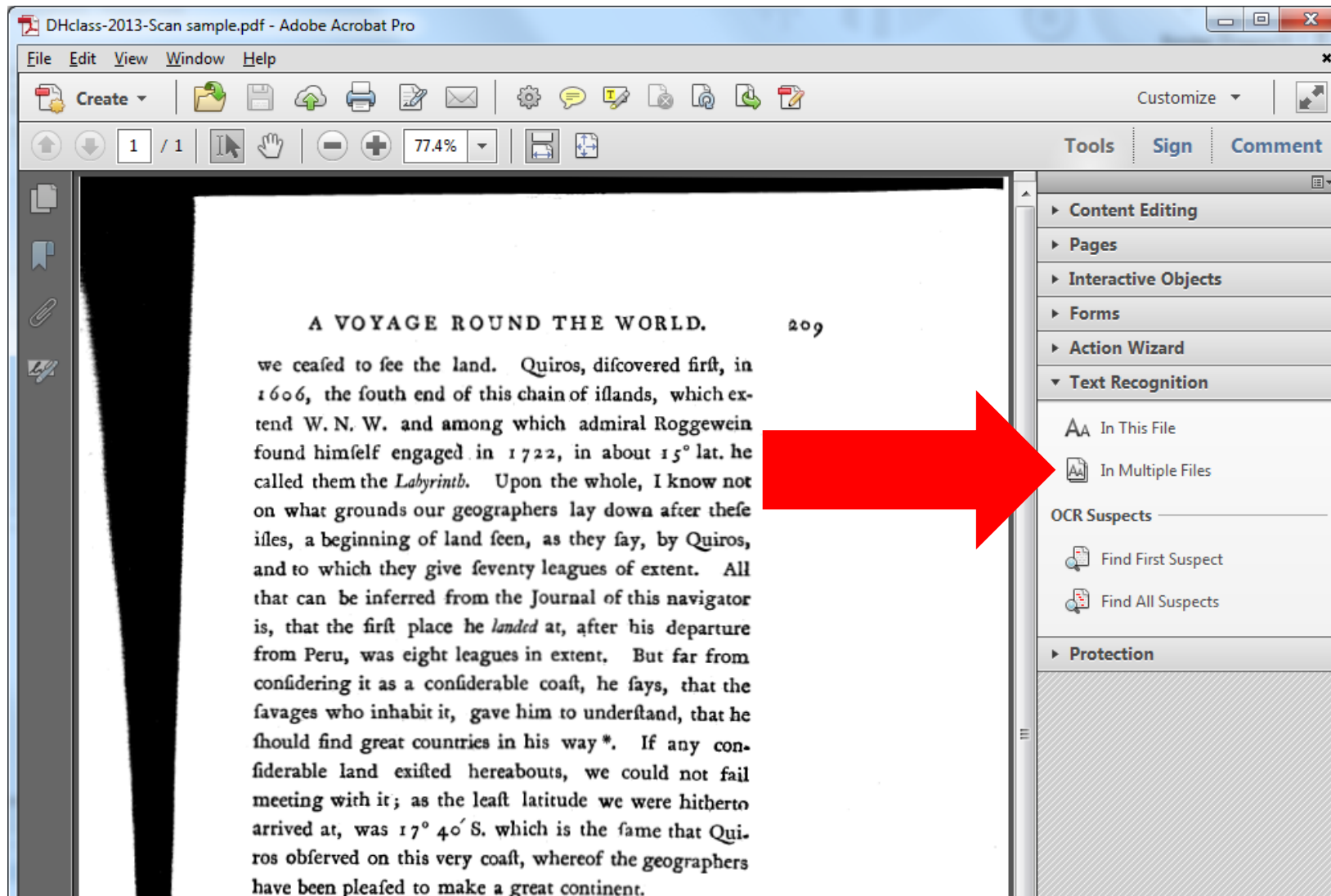




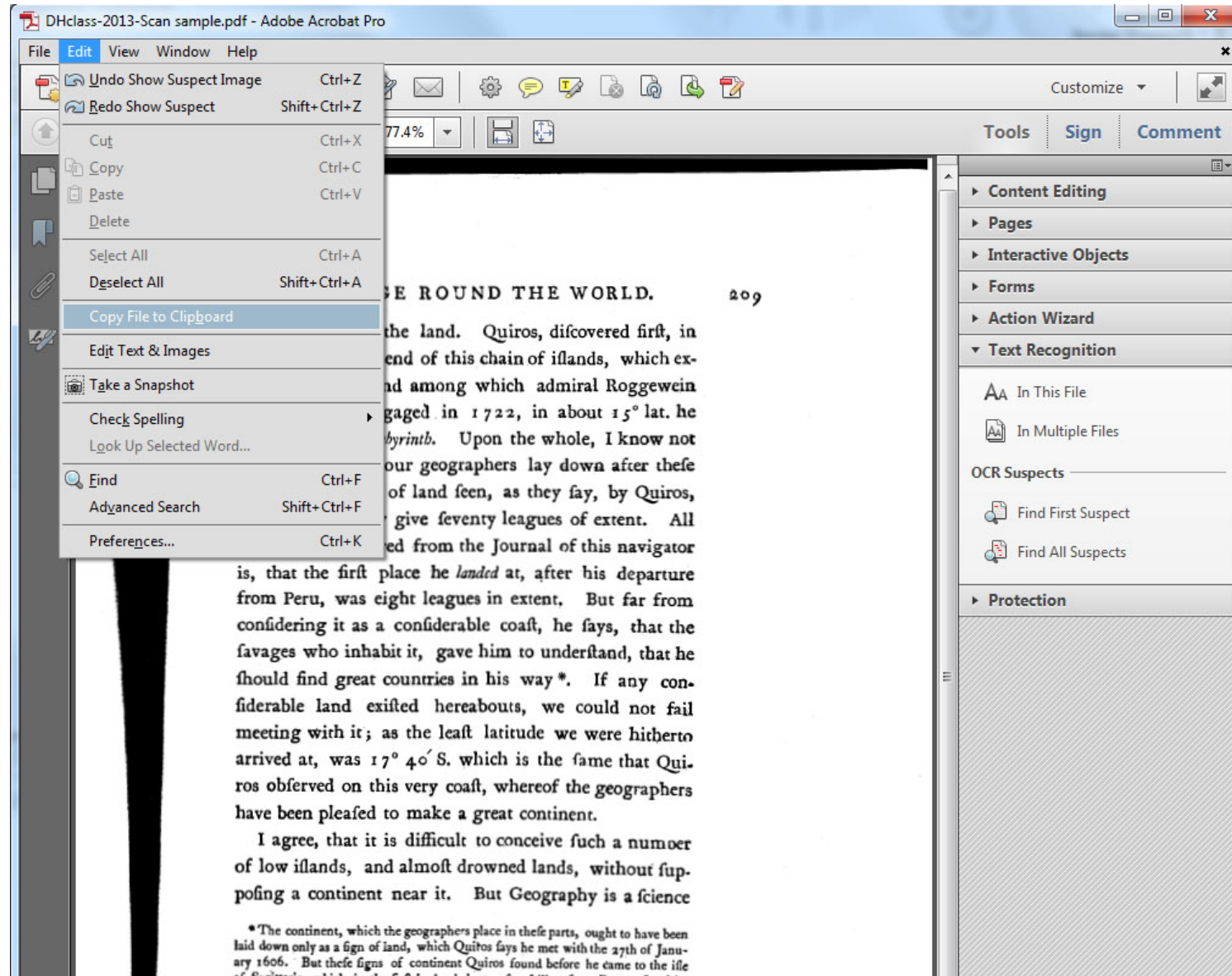
8) The document now includes formatted text, though probably with many errors.



- 9) Under Text Recognition, select Find All Suspects
- The words in red are the words that Acrobat SUSPECTS might be incorrect.



# 10) Under the Edit menu, select Copy File to Clipboard



11) ...and paste into a New Document in MS-Word to see how Adobe Acrobat Pro actually read the PDF image

12) Repeat for the second PDF image file that you downloaded

- In a brief one-page response paper due before class on Monday, Sept 9:
  - Discuss at least FIVE characteristics of the original text or PDF image documents that produced errors in the OCR process.
  - Include the OCR'd text from your MS-Word files as a separate pages with your response paper.
  - Please always remember to include your name in your submitted filename.