Visual Media Lab Research Outline

**Project**: This research is focused on empirically analyzing the visual tropes and conventions that are defined in the literature as stereotypes. Stereotypes are often the subject of popular and academic research, but as of now, case studies remain the only corpus analyzed, and this project seeks to expand the scope in order to report generalizable findings. This project will draw its sample from the Pulitzer Prizes.

**Objective:** Create a data set of images and metadata text from the [Pulitzer Prize Archives](https://www.pulitzer.org/prize-winners-by-category/216) from these three categories:

* Feature Photography 1995-2024.
* Breaking News Photography 2000-2024.
* Spot News Photography 1985-1999

Procedures:

1. Downloaded each image from every award-winning entry and create a unique ID (filename) for each image. Unique file names can follow this convention pp-sequence, e.g. pp-001, pp-002, pp-003). The images can be saved into this Onedrive folder here:

[Pulitzer Prizes](https://iowa-my.sharepoint.com/:f:/g/personal/bingbzhang_uiowa_edu/Eh0YUqtZO5tJqBxjbHispykB0QfN4wqLjRa0vkcSOGz2Pw?e=52FY91)

1. The unique file name should correspond to a row on a spreadsheet that contains the following metadata that should be scraped for each image:

* Date Awarded (Year only)
* Photographer Name
* Photographer’s country of origin (if possible)
* If Group Award, what group (e.g., Staff of Reuters, Staff of Associated Press)
* Location of where image was taken (city and country or just city).
* News Outlet worked for (if applicable, e.g. Dallas Morning News, Associated Press, etc)
* Competition Category (e.g. Spot News, Breaking News, Feature Photography)
* Caption

Data can be stored here:

<https://iowa-my.sharepoint.com/:x:/r/personal/bingbzhang_uiowa_edu/_layouts/15/Doc.aspx?sourcedoc=%7B1B8A399A-F20B-4132-AAD5-98B77B1ABBDF%7D&file=Coding%20Sheet%20for%20Award%20Winning%20Photos.xlsx&wdLOR=c991A81DF-9270-B44F-B4FB-1BC89BD6B8C1&action=default&mobileredirect=true>

It is possible that the images can be collected first, named, and then a secondary scraping procedure could create the spreadsheet corresponding to the images. Whatever workflow works for you, works for us.

Please reach out to me with any questions or concerns at any part of the process!