**R-ladies Outline**

**Data**: Nazi Era Provenance Internet Portal dataset for the NGA

**From Laurel RE: Art Data: “**NEPIP contains items from 179 museums that have listed artworks that they consider to have gaps in the history of ownership during the Nazi era.

The items in NEPIP\_NGA were selected from **one museum: The National Gallery of Art** in Washington DC.

Based on the identifier ("accession number") for the item in NEPIP the provenance (ownership history) of the item has been retrieved from the NGA's public collections site.

Thus, NEPIP\_NGA should contain the provenance (ownership history) that the NGA published online for every artwork that it posted on NEPIP, over 7000 items.

(It does not contain items that the NGA did not list on NEPIP.

**GOAL: Use R to rank the items by the probability that there is a serious problem with the provenance. Answer the question: Where should rare and expensive (human) expert research resources be focused?**

**Proposed Method: (At first) use lists of "Red Flags" that have been composed by humans based on a variety of sources. Then, when that works well, see if one can use machine learning to detect patterns that the humans may have missed.**

Red Flags: There are several lists of Red Flags of different types. The first and most official one is the 1946 Office of Strategic Services Art Looting Investigation Unit Final Report which lists 1000 names of people investigators considered to have been involved in the Nazi looted art market.

The list of names in its raw form can be found here:

[https://www.lootedart.com/MVI3RM469661](https://nam01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.lootedart.com%2FMVI3RM469661&data=02%7C01%7Clayla.bouzoubaa%40med.miami.edu%7C7d54fcdb800348bed47008d69aa9ef1c%7C2a144b72f23942d48c0e6f0f17c48e33%7C0%7C0%7C636866451988229577&sdata=MaDxQ1un3He9wc7DCV2hb8f7BSiN14AjAIrjMhscyh4%3D&reserved=0)

(I have a list of last names ready to use).

There are, in addition, other potential lists of Red Flag. These include:

- list of suspicious words or phrases in provenance

- list of people known to have been involved in trafficking but not on the 1946 list

- groupings of certain sellers”

**Workshop Outline:**

1. Version control & workflow in RStudio
   1. Objectives of the next 6 months
      1. Be able to work on a project from start to finish and have something to show for it
      2. collaborate
   2. Present art data
   3. Version control using git & github with Rstudio
   4. Workflow efficiencies in R (usethis and here packages)
      1. <https://jennybc.github.io/wtf-2019-rsc/>
2. Reproducible research/publication ready docs in RMarkdown
   1. <https://bookdown.org/yihui/rmarkdown/>
   2. Keep track of your steps
   3. Formatting section headings, chapters, TOC, font size
   4. Referencing tables, figures, chapters
   5. Inserting images, figures, tables
3. Data wrangling in the tidyverse / descriptives
   1. Debugging here??
4. Data visualization with ggplot2
5. Modeling/ machine learning
6. Building a Shiny app