## Managing and verifying data in HUGGO datasets

Two variables, ‘Checked_HUGGO’ and ‘Confirmed_HUGGO’ have been added to track progress on manually correcting entries.

‘Checked_HUGGO’: code 1 when the row's Text and Date (Signature, Force, End) observations have been verified and updated
‘Confirmed_HUGGO’: list variables for which the observation could be verified and confirmed.

- Eg. List 'Signature' in `Confirmed_HUGGO` if the Signature date was found and verified in the treaty text or in a manual online search.
- If TreatyText, MainText, AppendixText, or AnnexText are NAs, you can list them in ‘Confirmed_HUGGO’ if a manual search online does not return any texts.

## Cleaning and verifying treaty texts

The preface, preamble, articles, appendices and annexes should be retained.

For large pdf files (eg. GATT, EU treaties): Convert pdf file to word document, remove table of contents and headings or page numbers. Copy and paste the full text into the ‘TreatyText’ variable, the main text including the preamble and preface into ‘MainText’, the appendices in ‘AppendixText’, and the annexes in ‘AnnexText’. Make a note in the preparation script that this agreement was downloaded, converted, then added into R manually. Save the converted texts in Google drive or Dropbox for now.

For shorter texts or texts scraped from web pages: Remove in R any extra weblinks, javascript tags, strings of characters and symbols (eg. @, #, $, %, < >) that are not part of the treaty text. Include in the preparation script the code you used.

‘Members’ variable: list signatories found in the treaty text (separate states’ names with a comma), if available.