

Introducing palletjack:

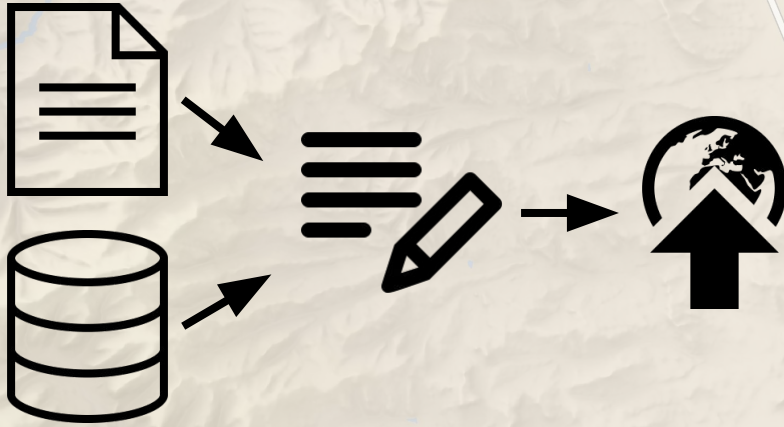
A Python library for updating AGOL
Hosted Feature Service Layers from
pandas DataFrames

Jacob Adams
UGIC 2023



Woodruff

ETL: How to Load Data



Woodruff

Explain Extract, Transform, and Load steps

Hitting the Hurdles

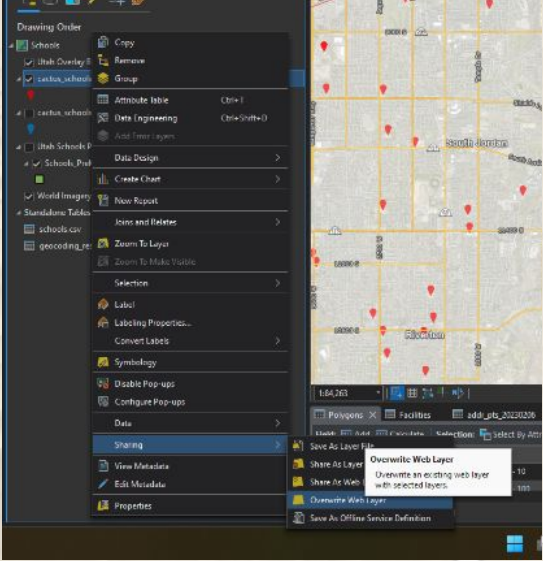


Credit: Kate Berg @pokatec_

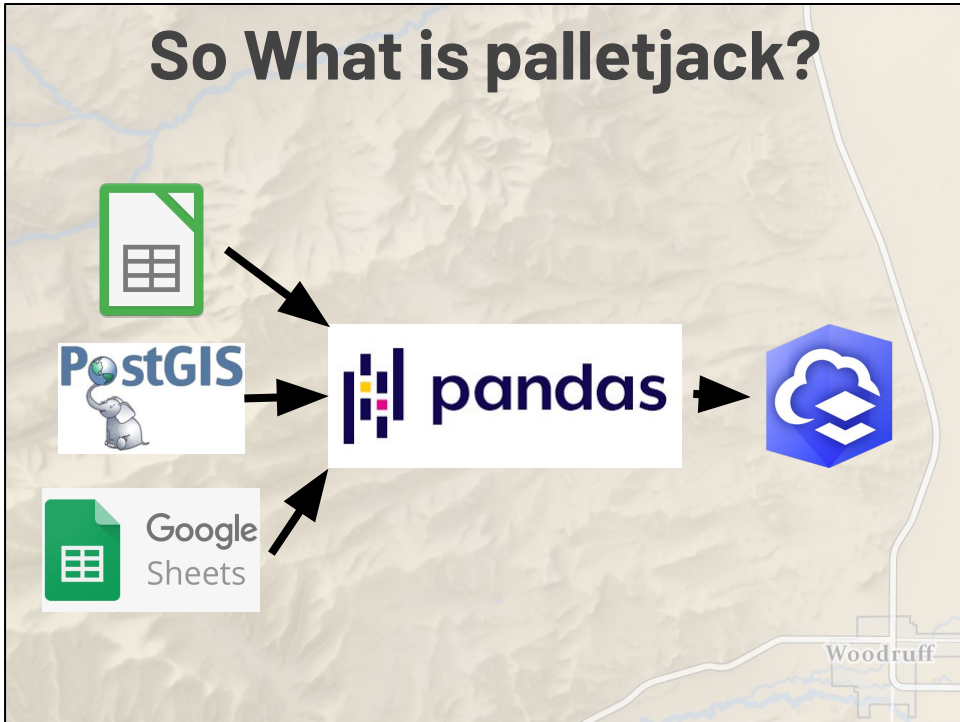
Woodruff

- Many different data sources and formats
- Data munging, joining, and cleaning
- Automation
- Reporting
- Reinventing the wheel

Doing it the Hard Way



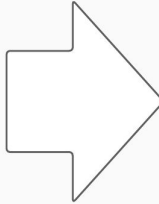
So What is palletjack?



- A python library of classes and methods for extracting data from various sources and then loading it into ArcGIS Online
- Uses the ArcGIS API for Python, no Pro/Enterprise needed- can run locally, on a server, or in the cloud

How Does it Help Me?

Re-inventing the Wheel
Every Time



Using palletjack for
Common Operations



- palletjack does all the heavy lifting of extracting and loading, allowing you to focus on your project-specific transforms

Show Me.

Utah Recycling and Solid Waste Facility Locator

Recycling and Solid Waste Facilities

- Compost or Other Food Management Facility
- Incinerator
- Land Treatment or Landspreading Facility
- Landfill
- Recycling Facility
- Transfer Station
- Used Oil Collection Center (UOCC)
- Waste Tire Facility

Glossary
MSW: Municipal Solid Waste
CD: Construction and Demolition
[Click here](#) for detailed descriptions of the various facility types.

Lehi CD Landfill
Landfill - Class IVb
2450 North N 300 West
Lehi, 84043 UT
Lat, Long: [40.41667, -111.85](#)
Website:
Phone: (801) 768-7102
Tons of Material Diverted from Landfills Last Year:
Not Yet Reported
Last Updated: 4 Apr 2023

Filter By County
County Name: - All -

Filter by Facility Type
Facility Type: - All -

Accepts Material from the Public (Contact the facility for information)

Search for Certified Electronics Recyclers:

- e-Stewards
- R2

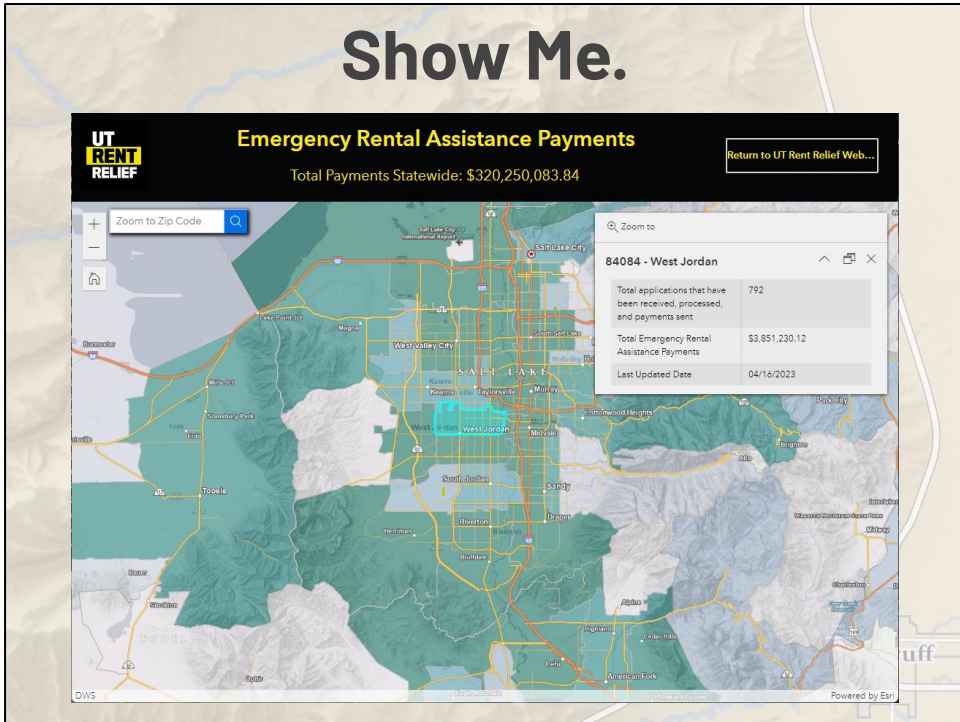
Search for recycling facilities by material type:

- Earth911

Powered by Esri

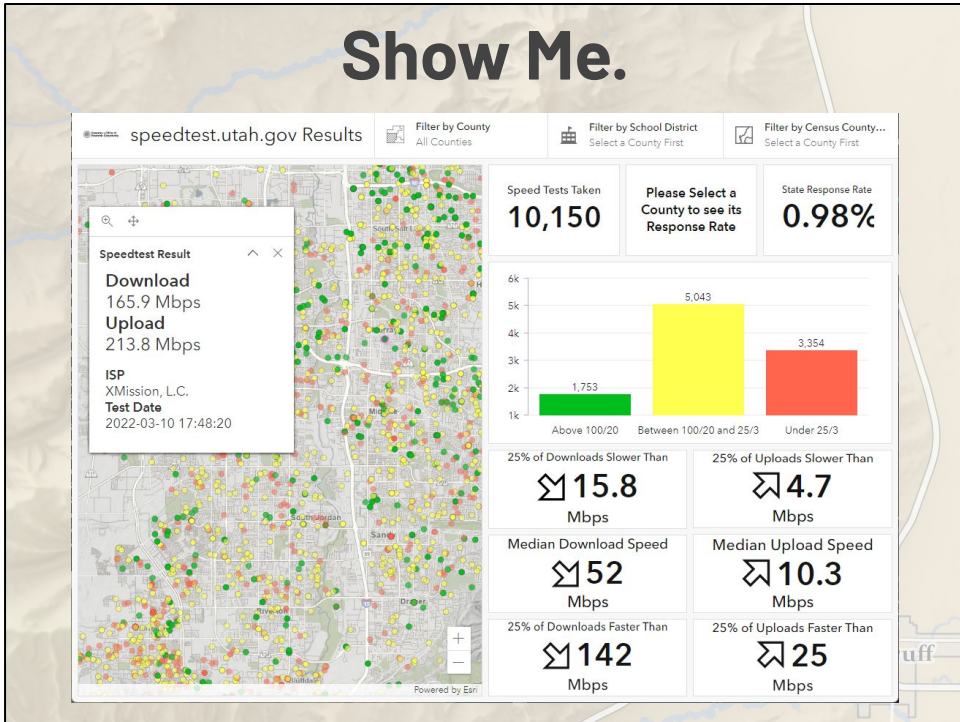
- ERAP
- DWMRC
- Speedtest

Show Me.



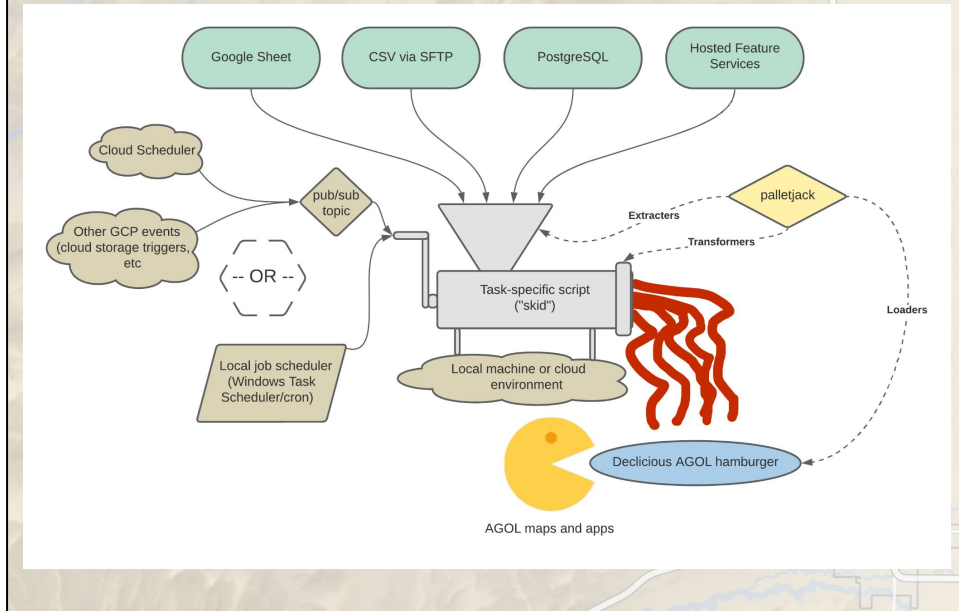
- ERAP
- DWMRC
- Speedtest

Show Me.



- ERAP
- DWMRC
- Speedtest

How Does palletjack Work?



- The palletjack meatgrinder
- Users write their own scripts using palletjack classes and methods
- The script extracts data from your sources, transforms it, and loads it into the hosted feature service
- The script lives on either a local machine or runs in the cloud
- Gets kicked off manually or through some scheduling tool

Pandas at the Core

	hex_id	SHAPE
0	8928b654e03ffff	{"rings": [[[-111.79321259099999, 41.827124674...
1	8929925ab57ffff	{"rings": [[[-110.38265724499996, 37.936864283...
2	892693aee0bffff	{"rings": [[[-110.38752944799995, 38.980448737...
3	89299e18987ffff	{"rings": [[[-113.42266556199996, 39.046304283...
4	8928b4b0a2ffff	{"rings": [[[-114.05179814599995, 41.089726935...
...
1941327	89299a49377ffff	{"rings": [[[-112.67134165099998, 40.368783260...
1941328	892692b14d3ffff	{"rings": [[[-111.75507706899998, 39.872849589...
1941329	892992c608bffff	{"rings": [[[-110.89404863599998, 37.634269653...
1941330	89269148667ffff	{"rings": [[[-110.43642184999999, 39.436238886...
1941331	8928b6c0337ffff	{"rings": [[[-111.79206106199996, 41.512871236...

1941332 rows × 2 columns

- palletjack uses pandas DataFrames as the common data format between each of the ETL steps
- Spatial data handled via ArcGIS API for Python's Spatially-Enabled DataFrames
- Extract data from various sources as dataframes
- Transform, clean, and merge into a single dataframe
- Load final dataframe into AGOL

Architecting in AGOL

Service List

ArcGIS REST Services Directory

Home > services

2020

Folder: /

Current Version: 11

Services

- 2017 Outdoor Recreation Grant Overview (FeatureServer)
- 2018 Outdoor Recreation Grant Overlay (FeatureServer)
- 2019 Outdoor Recreation Grant (FeatureServer)
- 2011 DriveCast (FeatureServer)
- 2016 Utah Lidar Acquisition Areas (FeatureServer)
- 2020 Outdoor Recreation Grant (FeatureServer)
- 2021 Outdoor Recreation Grant (FeatureServer)
- 2021 Qualified Census Tracts (FeatureServer)
- Access to Parks (FeatureServer)
- ACS Housing Tracts (FeatureServer)
- ACS Indicators Cleaned (FeatureServer)
- Address Points Grand San Juan (FeatureServer)
- Aerial Photography Exports (FeatureServer)
- AGOL Upload (FeatureServer)
- Airport Locations (FeatureServer)
- Alma 100 20 Outlets (FeatureServer)
- AlisFvetsGeobata_points (FeatureServer)
- Alluvial Fans (FeatureServer)
- Ambulance Service Areas (FeatureServer)
- Aquifer Basin Fill Boundary (FeatureServer)
- Aquifer Recharge Discharge Areas (FeatureServer)
- Attachment Text (FeatureServer)

Feature Service

ArcGIS REST Services Directory

Home > services > UtahCountyBoundaries (FeatureServer)

2020

UtahCountyBoundaries (FeatureServer)

View In: [Map Viewer](#)

Service Description: This data set represents county boundaries.

Service ItemId: 90431cac29f49f4bcf1505419583753

Has Versioned Data: false

Max Record Count: 2000

Supported query Formats: JSON

Supports applyEdits with GlobalIds: True

[All Layers and Tables](#)

Layers

- Counties (0)

Description:

This dataset reflects the most current version of Utah county boundaries agreement (certified 20120612) between Juab and Millard Counties, Any BLM (CadNSDIV2), any where from 10 to 50 feet all in non-populated areas. Data is current through Nov. 15, 2021; Population Figures from 2020 Ce

Feature Service Layer

ArcGIS REST Services Directory

Home > services > UtahCountyBoundaries (FeatureServer) > Counties

2020

Layer: Counties (ID:0)

View In: [Map Viewer](#)

Name: Counties

Display Field: NAME

Type: Feature Layer

Geometry Type: esriGeometryPolygon

Description:

This dataset reflects the most current version of Utah county boundaries agreement (certified 20120612) between Juab and Millard Counties, Any BLM (CadNSDIV2), any where from 10 to 50 feet all in non-populated areas. Data is current through Nov. 15, 2021; Population Figures from 2020 Ce

Copyright Text:

Min. Scale: 0

Max. Scale: 0

Default Visibility: true

Max Record Count: 2000

Supported query Formats: JSON

Use Standardized Queries: True

- palletjack operates on layers within a Feature Service

How Do I Use it?

WHAT GIVES PEOPLE FEELINGS OF POWER



Woodruff

Four Easy Steps to Use palletjack

1. Use an extractor in the `extract` module, or any other method, to load your data into a pandas DataFrame
2. Merge your new data with any existing data that you want to save or update, or prepare your new data to completely overwrite the existing data
3. Clean the new data to align with the right column names and data types for AGOL using the `transform` and `utils` modules
4. Add new features, overwrite existing features, and/or delete out-of-date features in AGOL with the `load` module

- Not a stand-alone program, but classes and methods to be used in your own scripts
- `pip install ugrc-palletjack`
- Each step is its own module with in `palletjack`
- `from palletjack import extract, load, transform, utils`

palletjack in PyPI



pypi.org/project/ugrc-palletjack

```
C:\Users\jdadams  
(palletjack) λ pip install -U ugrc-palletjack
```

- Not a stand-alone program, but classes and methods to be used in your own scripts
- pip install ugrc-palletjack

Modules: palletjack's Pieces

extract

transform

utils

load

Woodruff

- Each step is its own module within palletjack
- `from palletjack import extract, load, transform, utils`

extract: Getting Data



Google
Sheets



Classes

GSheetLoader

```
combine_worksheets_into_single_dataframe  
load_all_worksheets_into_dataframes  
load_specific_worksheet_into_dataframe
```

GoogleDriveDownloader

```
download_attachments_from_dataframe  
download_attachments_from_dataframe_using_a  
pi  
download_file_from_google_drive  
download_file_from_google_drive_using_api
```

PostgresLoader

```
read_table_into_dataframe
```

SFTPLoader

```
download_sftp_folder_contents  
download_sftp_single_file  
read_csv_into_dataframe
```

- palletjack.extract
- other, simpler ways of loading data- pd.read_csv, etc

transform and utils

Classes

APIGeocoder

`geocode_dataframe`

DataCleaning

`rename_dataframe_columns_for_agol`

`switch_to_datetime`

`switch_to_float`

`switch_to_nullable_int`

FeatureServiceMerging

`get_live_dataframe`

`update_live_data_with_new_data`

- palletjack's purpose is to let you focus your energy on this step, not the common and recurring problems of extracting data or loading it into AGOL
- A few helper methods in `palletjack.transform` for merging data, including the `DataCleaning` class
- Geocoding in `palletjack.utils`
- `retry` in `palletjack.utils`

transform.DataCleaning

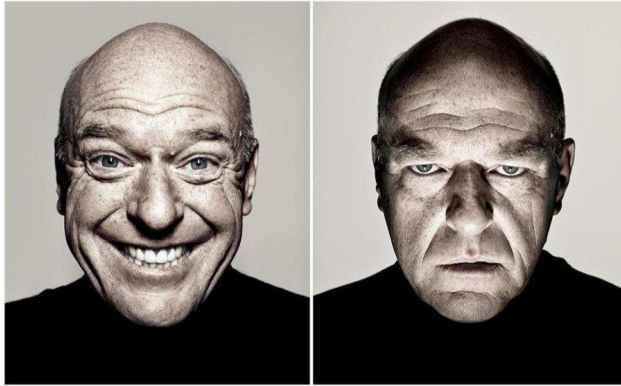


- The feature service updater will raise warnings and fail if the data aren't in perfect alignment.
- Compile list of things to check for and fix
- Show how DataCleaning class can help

transform.DataCleaning

"Yeah, I have the data for your map already in a spreadsheet!"

spreadsheet has 10 sheets with formulas, merged cells, and blank rows between data.



Credit: Kate Berg @pokateo_

- Your data are unique, and so are their problems.
- Some things you just have to sort out yourself
- Pandas makes this simpler and easier than a bunch of arcpy GP tools
- MS Data Wrangler for VSCode

load: to AGOL, to AGOL!

Classes

ColorRampReclassifier

`update_color_ramp_values`

FeatureServiceAttachmentsUpdater

`build_attachments_dataframe`

`update_attachments`

FeatureServiceUpdater

`add_features`

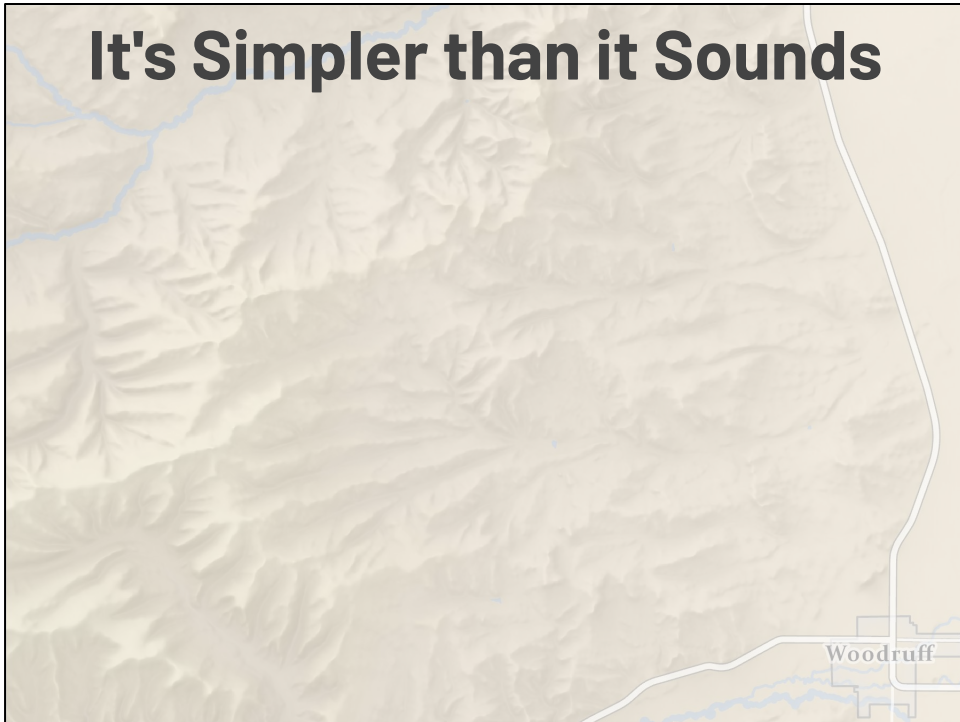
`remove_features`

`truncate_and_load_features`

`update_features`

- FeatureServiceUpdater does all the lifting
- Also an attachment updater and a very simple map color reclassifier (that needs to be redone/rewritten)
- Uses class methods so you don't have to create the object
- Four methods, each handling one specific task. It's your job to divide up your incoming data into the appropriate steps
- If you're getting all the new data as part of your load step, it may be easier to just truncate and load.

It's Simpler than it Sounds



- Show code for WMRC, speedtest skids

Help Me!

Read the Documentation:

agrc.github.io/palletjack/palletjack/

Check the Issue Tracker:

<https://github.com/agrc/palletjack/issues>

Crib Code from Our Skids:

**[https://github.com/search?q=org%3Aagrc
+skid&type=repositories](https://github.com/search?q=org%3Aagrc%20skid&type=repositories)**

Woodruff



jdadams@utah.gov
gis.utah.gov/presentations

