



tidyagronomy

An extensible framework & ecosystem for High-Speed
Agronomic Feature Generation, At-Scale

James Horine



Special thanks

- osQF organizers
- Barchart & Cody Gerlach
- Tim Krause & Kate Zipp
- The audience here today



Overview

- [tidyagronomy](#) is an [R package](#) that provides a scalable, tidyverse-based system to compute growing degree units (GDUs), stress degree units, and chilling units for thousands of locations efficiently.
- Designed for extensibility, it transforms high-frequency weather data into standardized, tidy data structures, enabling fast and reproducible agronomic modeling at regional to national scales.
- The framework empowers researchers to accelerate studies in crop growth, phenology, pest modeling, soil interactions, and forecasting by making large-scale weather and agronomy analytics accessible and modular.



Tidyagronomy Workflow 0.1

- Web-search: Bring NetCDF weather file from ECMWF/ERA5
- Load/Extract from file into R session through tidync |>
- Convert brick to hypertibble |>
- Run appropriate aggregations for desired crops, pests, and conditions |>
- Utilize standard data science/modeling workflow in tidyverse |>
- mutate(something=...) |>
- profit(?)



tidyagronomy::data-scheme(a)

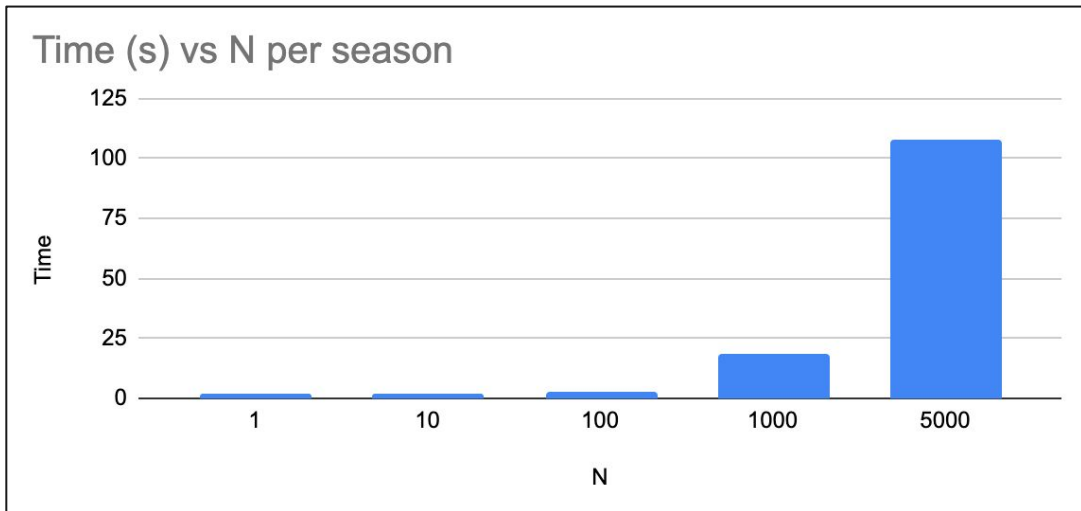
- The underlying packaging requires time series that are time AND space aligned!
- Efficient computation/declaration of location identifiers by the use of geohashes
 - (+) easy for the user to understand
 - (sort of) String lookups are fast in memory
 - (good enough) and reliable to use as a join key

Inherent or natural dimensions of data brick/tensor				Defined/Computed feature(s)		
latitude	longitude	observation	timestamp	f(observation)	geohash(7)	season
50	-104	273.15K	2016-04-01 00:00:00	0	c8yh3qj	2016

Preliminary Code Profiling

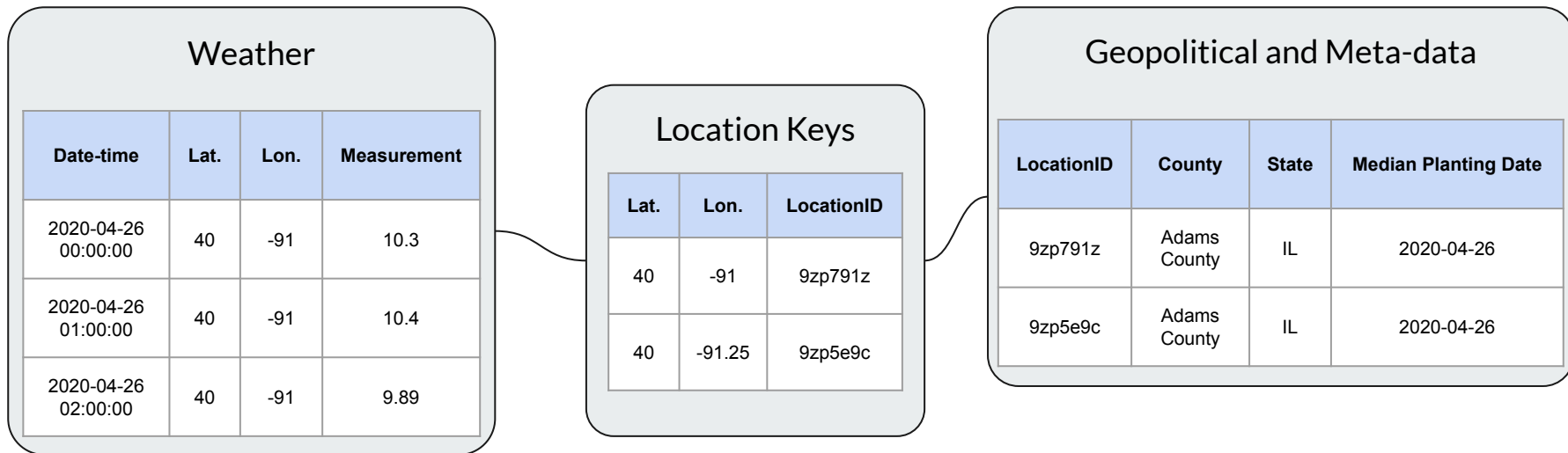
- For hundreds of location-seasons (~ 900) the compute time is on the order of single seconds.
- For thousands... <20 seconds
- For 5000(ish) < 115 seconds

(on an aging macbook air)

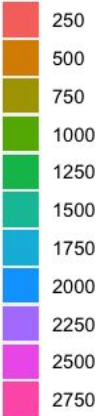
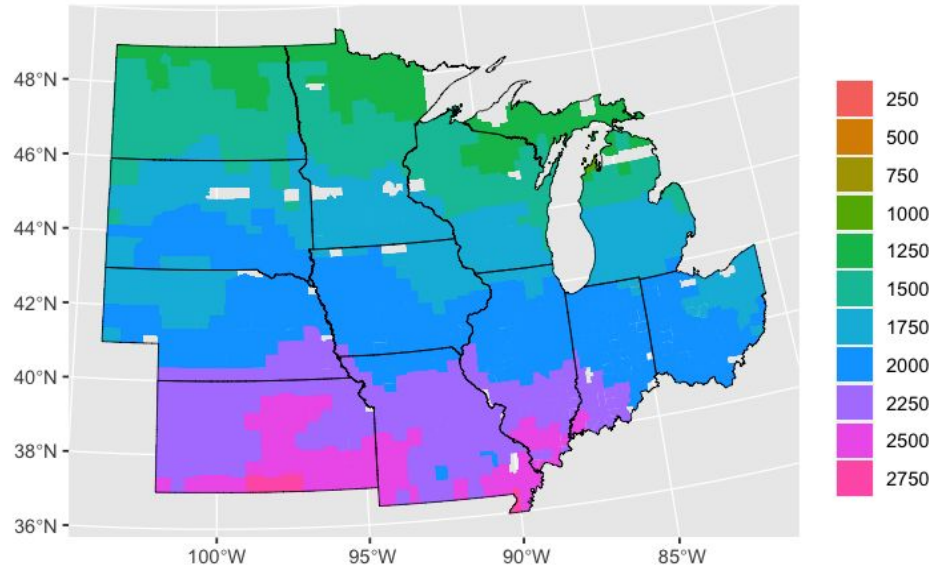
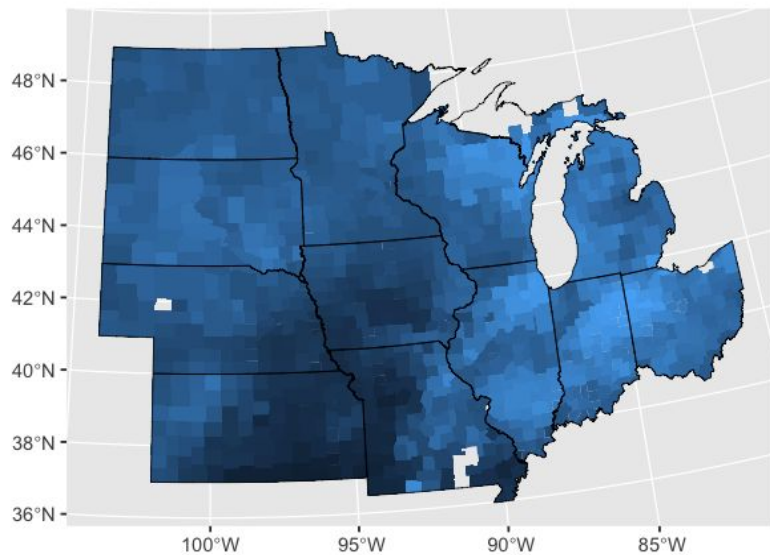




Visualizing the join structure in TidyAgronomy

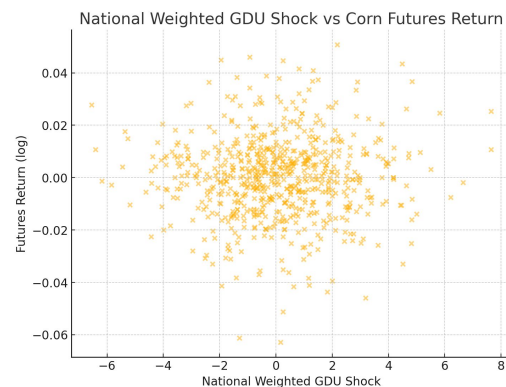
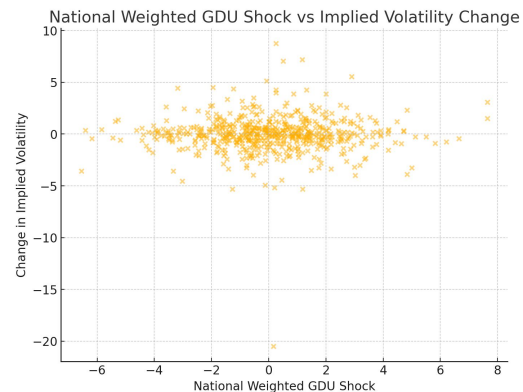


Median Planting Date VS GDU Potential



Sample Market Impacts

Model	Term	Coefficient	Std. Error	t-Stat	p-Value
Futures Return (Lag 0)	GDU Shock	0.000117	0.000301	0.39	0.698
Futures Return (Lag 3)	GDU Shock	0.000354	0.000337	1.05	0.293
Futures Return (Lag 7)	GDU Shock	0.000363	0.000339	1.07	0.285
Futures Return (Lag 10)	GDU Shock	-0.000144	0.000337	-0.43	0.670
Quadratic Model	GDU Shock	0.000072	0.000282	0.25	0.800
Quadratic Model	GDU Shock ²	0.000145	0.000084	1.73	0.084





Links

- [Manuscript Draft](#)
- [Repo](#)
- [James Horine](#)