Data Science, Time Complexity & Spacekime Analytics

Ivo D. Dinov

Statistics Online Computational Resource
Health Behavior & Biological Sciences
Computational Medicine & Bioinformatics
Neuroscience Graduate Program
Michigan Institute for Data Science
University of Michigan

https://SOCR.umich.edu

Joint work with Milen V. Velev (BTU)

Based on an upcoming book "*Data Science: Time Complexity, Inferential Uncertainty & Spacekime Analytics*"

SCHOOL OF NURSING STATISTICS ONLINE COMPUTATIONAL RESOURCE (SOCR) UNIVERSITY OF MICHIGAN

Slides Online: "SOCR News"

Outline

- Motivation: Big Data Analytics Challenges
- Complex-Time (*kime*)
- Spacekime Calculus & Math Foundations
- Open Spacekime Problems
- Neuroscience Applications
  - Longitudinal Neuroimaging (UKBB, fMRI)

Big Data Characteristics & Challenges

IBM Big Data 4V's: Volume, Variety, Velocity & Veracity

Big Bio Data Dimensions	Specific Challenges
Size	Harvesting and management of vast amounts of data
Complexity	Wranglers for dealing with heterogeneous data
Incongruency	Tools for data harmonization and aggregation
Multi-source	Transfer, joint multivariate representation & modeling
Multi-scale	Interpreting macro → meso → micro → nano scale observations
Time	Techniques accounting for longitudinal effects (e.g., time corr)
Incomplete	Reliable management of missing data, imputation, obfuscation

Example: analyzing observational data of 1,000's Parkinson's disease patients based on 10,000's signature biomarkers derived from multi-source imaging, genetics, clinical, physiologic, phenomics and demographic data elements

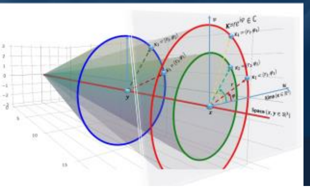
Software developments, student training, service platforms and methodological advances associated with the Big Data Discovery Science all present existing opportunities for learners, educators, researchers, practitioners and policy makers

Dinov, *GigaScience* (2016)

Gao *et al., SciRep* (2018)

Complex-Time (*Kime*)

- At a given spatial location, $\boldsymbol{x}$, complex time (*kime*) is defined by $\kappa = re^{i\varphi} \in \mathbb{C}$, where:
  - the magnitude represents the longitudinal events order ($r > 0$) and characterizes the longitudinal displacement in time, and
  - event phase ($-\pi \le \varphi < \pi$) is an angular displacement, or event direction
- There are multiple alternative parametrizations of kime in the complex plane
- Space-kime manifold is $\mathbb{R}^3 \times \mathbb{C}$:
  - $(\boldsymbol{x}, k_1)$ and $(\boldsymbol{x}, k_4)$ have the same spacetime representation, but different spacekime coordinates,
  - $(\boldsymbol{x}, k_1)$ and $(\boldsymbol{y}, k_1)$ share the same kime, but represent different spatial locations,
  - $(\boldsymbol{x}, k_2)$ and $(\boldsymbol{x}, k_3)$ have the same spatial-locations and kime-directions, but appear sequentially in order, $r_2 < r_1$.



Rationale for *Time* → *Kime* Extension

- **Math**: $Time$ is a special case of *kime*, $\kappa = |\kappa|e^{i\varphi}$ where $\varphi = 0$ (nil-phase)
  - algebraically a *multiplicative* (algebraic) group, (multiplicative) unity (identity) $= 1$
  - multiplicative inverses, multiplicative identity, associativity $t_1 * (t_2 * t_3) = (t_1 * t_2) * t_3$
  - $time$ is **not** a complete *algebraic field* $(+,*)$:
    - Additive unity (0), element additive inverse $(-t)$: $t + (-t) = 0$; is outside $\mathbb{R}^+$ (time-domain)
    - $x^2 + 1 = 0$ has no solutions in time (or in $\mathbb{R}$) ....

$$\text{Group}(*) \subseteq Ring\left(\underbrace{\overbrace{(+,*)}^{\text{Compatible operations}}}_{\text{associative \& distributive}}\right) \subseteq Field\ \overbrace{(+,*)}^{\text{Group}(+)}$$

  - Classical time ($\mathbb{R}^+$) is a *positive cone* over the field of the real numbers ($\mathbb{R}$)
  - Time forms a subgroup of the multiplicative group of the reals
  - Whereas kime ($\mathbb{C}$) is an algebraically *closed prime field* that naturally extends time
  - *Time* is ordered & *kime* is not – the kime magnitude preserves the intrinsic time order
  - Kime ($\mathbb{C}$) represents the smallest natural extension of time, complete filed that agrees with time
  - The $time *$ group is closed under addition, multiplication, and division (but not subtraction). It has the topology of $\mathbb{R}$ and the structure of a multiplicative topological group ≡ additive topological semigroup
- **Physics** – Problems of time … (DOI 10.1007/978-3-319-58848-3)
- **AI/Data Science** – Random IID sampling, Bayesian reps, tensor modeling of $\mathbb{C}$ kimesurfaces, novel analytics

Dinov & Velev (2021)

The Spacekime Manifold

- **Spacekime:** $(\boldsymbol{x}, \boldsymbol{k}) = \left(\underbrace{x^1, x^2, x^3}_{\text{Point in space}}, \underbrace{c\kappa_1 = x^4, c\kappa_2 = x^5}_{\text{Moment in kime}}\right) \in X, \quad c \sim 3 \times 10^8\ m/s$
- **Kevents** (*complex events*): points (or states) in the spacekime manifold X. Each kevent is defined by where ($\boldsymbol{x} = (x, y, z)$) it occurs in space, what is its *causal longitudinal order* $\left(r = \sqrt{(x^4)^2 + (x^5)^2}\right)$, and in what *kime-direction* ($\varphi = \text{atan2}(x^5, x^4)$) it takes place.
- **Spacekime interval** (ds) is defined using the general Minkowski 5×5 metric tensor $(\lambda_{ij})_{i=1,j=1}^{5,5}$, which characterizes the geometry of the *(generally curved) spacekime* manifold:

$$ds^2 = \sum_{i=1}^{5}\sum_{j=1}^{5} \lambda_{ij} dx^i dx^j = \lambda_{ij} dx^i dx^j$$

- *Euclidean (flat) spacekime* metric corresponds to the tensor:

$$(\lambda_{ij}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

  - Spacelike intervals correspond to $ds^2 > 0$, where an inertial frame can be found such that two kevents $a, b \in X$ are simultaneous. An object can't be present at two kevents which are separated by a spacelike interval.
  - Lightlike intervals correspond to $ds^2 = 0$. If two kevents are on the line of a photon, then they are separated by a lightlike interval and a ray of light could travel between the two kevents.
  - Kimelike intervals correspond to $ds^2 < 0$. An object can be present at two different kevents, which are separated by a kimelike interval.

Spacekime Calculus

- Kime **Wirtinger derivative** (first order kime-derivative at $\boldsymbol{k} = (r, \varphi)$):

$$f'(z) = \frac{\partial f(z)}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right) \text{ and } f'(\bar z) = \frac{\partial f(\bar z)}{\partial \bar z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right).$$

In Conjugate-pair basis: $df = \partial f + \bar\partial f = \frac{\partial f}{\partial z}dz + \frac{\partial}{\partial \bar z}d\bar z$

In Polar kime coordinates:

$$f'(k) = \frac{\partial f(k)}{\partial k} = \frac{1}{2}\left(\cos\varphi\frac{\partial f}{\partial r} - \frac{1}{r}\sin\varphi\frac{\partial f}{\partial \varphi} - i\left(\sin\varphi\frac{\partial f}{\partial r} + \frac{1}{r}\cos\varphi\frac{\partial f}{\partial \varphi}\right)\right) = \frac{e^{-i\varphi}}{2}\left(\frac{\partial f}{\partial r} - \frac{i}{r}\frac{\partial f}{\partial \varphi}\right)$$

$$f'(\bar k) = \frac{\partial f(\bar k)}{\partial \bar k} = \frac{1}{2}\left(\cos\varphi\frac{\partial f}{\partial r} - \frac{1}{r}\sin\varphi\frac{\partial f}{\partial \varphi} + i\left(\sin\varphi\frac{\partial f}{\partial r} + \frac{1}{r}\cos\varphi\frac{\partial f}{\partial \varphi}\right)\right) = \frac{e^{i\varphi}}{2}\left(\frac{\partial f}{\partial r} + \frac{i}{r}\frac{\partial f}{\partial \varphi}\right).$$

- Kime **Wirtinger integration**:

Path-integral $\lim_{|z_{i+1}-z_i|\to 0}\sum_{i=1}^{n-1}(f(z_i)(z_{i+1}-z_i)) \cong \oint_{z_a}^{z_b} f(z_i)dz$.

Definite area integral: for $\Omega \subseteq \mathbb{C}$, $\int_\Omega f(z)dzd\bar z$.

Indefinite integral: $\int f(z)dzd\bar z$, $df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial \bar z}d\bar z$.

The *Laplacian* in terms of conjugate pair coordinates is $\Delta f = d^2 f = 4\frac{\partial f}{dz}\frac{\partial f}{d\bar z} = 4\frac{\partial f}{d\bar z}\frac{\partial f}{dz}$.

Dinov & Velev, De Gruyter (2021), *in press*.

Spacekime Generalizations

- Spacekime generalization of **Lorentz transform** between two reference frames, K & K':

(the interval ds is Lorentz transform invariant)

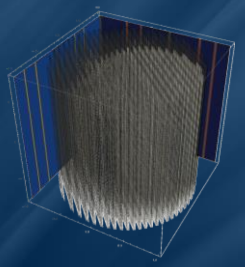
$$\underbrace{\begin{pmatrix} x' \\ y' \\ z' \\ k_1' \\ k_2' \end{pmatrix}}_{\in K'} = \begin{pmatrix} \zeta & 0 & 0 & -\frac{c^2}{v_1}\beta^2\zeta & -\frac{c^2}{v_2}\beta^2\zeta \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ -\frac{1}{v_1}\beta^2\zeta & 0 & 0 & 1+(\zeta-1)\frac{c^2}{(v_1)^2}\beta^2 & (\zeta-1)\frac{c^2}{v_1 v_2}\beta^2 \\ -\frac{1}{v_2}\beta^2\zeta & 0 & 0 & (\zeta-1)\frac{c^2}{v_1 v_2}\beta^2 & 1+(\zeta-1)\frac{c^2}{(v_2)^2}\beta^2 \end{pmatrix} \underbrace{\begin{pmatrix} x \\ y \\ z \\ k_1 \\ k_2 \end{pmatrix}}_{\in K}$$

where $0 \le \beta = \frac{1}{\sqrt{\left(\frac{c}{v_1}\right)^2 + \left(\frac{c}{v_2}\right)^2}} \le 1$ & $\zeta = \frac{1}{\sqrt{1-\beta^2}} \ge 1$.

Dinov & Velev (2021)

Spacekime Solution to Wave Equation

- Math Generalizations:
  Derived other spacekime concepts: law of addition of velocities, energy-momentum conservation law, stability conditions for particles moving in spacekime, conditions for nonzero rest particle mass, causal structure of spacekime, and solutions of the ultrahyperbolic wave equation under Cauchy initial data …



Wang et al., 2021 | Dinov & Velev (2021)

Ultrahyperbolic Wave Equation – Cauchy Initial Data

- **Nonlocal constraints** yield the existence, uniqueness & stability of local and global solutions to the ultrahyperbolic wave equation under Cauchy initial data …

$$\underbrace{\sum_{i=1}^{d_s}\partial_{x_i}^2 u \equiv \Delta_{\boldsymbol{x}} u(\boldsymbol{x},\boldsymbol{\kappa})}_{\text{spatial Laplacian}} = \underbrace{\Delta_{\boldsymbol{\kappa}} u(\boldsymbol{x},\boldsymbol{\kappa}) \equiv \sum_{i=1}^{d_t}\partial_{\kappa_i}^2 u}_{\text{temporal Laplacian}}, \quad \underbrace{\left| \begin{matrix} u_o = u\Big(\underbrace{\boldsymbol{x}}_{x\in D_s}, \underbrace{0,\boldsymbol{\kappa}_{-1}}_{\kappa\in D_t}\Big) = f(\boldsymbol{x},\boldsymbol{\kappa}_{-1}) \\ u_1 = \partial_{\kappa_1}u(\boldsymbol{x},0,\boldsymbol{\kappa}_{-1}) = g(\boldsymbol{x},\boldsymbol{\kappa}_{-1}) \end{matrix} \right.}_{\text{initial conditions (Cauchy Data)}}$$

where $\boldsymbol{x} = (x_1, x_2, \dots, x_{d_s}) \in \mathbb{R}^{d_s}$ and $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_{d_t}) \in \mathbb{R}^{d_t}$ are the Cartesian coordinates in the d_s space and d_t time dims.

Stable local solution over a Fourier frequency region defined by $|\boldsymbol{\xi}| \ge |\boldsymbol{\eta}_{-1}|$ … nonlocal constraints:

$$\hat u\Big(\boldsymbol{\xi}, \underbrace{\kappa_1, \boldsymbol{\eta}_{-1}}_{\eta}\Big) = \cos\left(2\pi\,\kappa_1\sqrt{|\boldsymbol{\xi}|^2 - |\boldsymbol{\eta}_{-1}|^2}\right)\underbrace{\hat u_o(\boldsymbol{\xi},\boldsymbol{\eta}_{-1})}_{c_1} + \sin\left(2\pi\,\kappa_1\sqrt{|\boldsymbol{\xi}|^2 - |\boldsymbol{\eta}_{-1}|^2}\right)\underbrace{\frac{\hat u_1(\boldsymbol{\xi},\boldsymbol{\eta}_{-1})}{2\pi\sqrt{|\boldsymbol{\xi}|^2 - |\boldsymbol{\eta}_{-1}|^2}}}_{c_2},$$

where $\mathcal{F}\begin{pmatrix} u_o \\ u_1 \end{pmatrix} = \begin{pmatrix} \hat u_o \\ \hat u_1 \end{pmatrix} = \begin{pmatrix} \hat u_o(\boldsymbol{\xi},\boldsymbol{\eta}_{-1}) \\ \hat u_1(\boldsymbol{\xi},\boldsymbol{\eta}_{-1}) \end{pmatrix} = \begin{pmatrix} \hat u(\boldsymbol{\xi},\boldsymbol{\eta}_{-1}) \\ \partial_{\kappa_1}\hat u(\boldsymbol{\xi},\boldsymbol{\eta}_{-1}) \end{pmatrix}$.

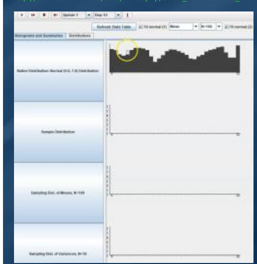
$$u\Big(\boldsymbol{x}, \underbrace{\kappa_1, \boldsymbol{\kappa}_{-1}}_{\kappa}\Big) = \mathcal{F}^{-1}(\hat u)(\boldsymbol{x},\boldsymbol{\kappa}) = \int_{\hat D_s \times \hat D_{t_{-1}}} \hat u(\boldsymbol{\xi}, \kappa_1, \boldsymbol{\eta}_{-1}) \times e^{2\pi i\langle x,\xi\rangle} \times e^{2\pi i\langle \kappa_{-1},\eta_{-1}\rangle} d\boldsymbol{\xi}\, d\boldsymbol{\eta}_{-1}.$$

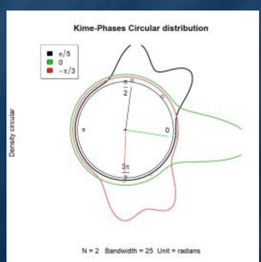
Wang et al., 2021 | Dinov & Velev (2021)

Hidden Variable Theory & Random Sampling

- Kime phase distributions are mostly symmetric, random observations ≡ phase sampling

http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_GeneralCentralLimitTheorem





https://www.socr.umich.edu/TCIU/HTMLs/Chapter6_Kime_Phases_Circular.html

Dinov, Christou & Sanchez (2008) Dinov & Velev (2021)

(Many) Spacekime Open Math Problems

- **Ergodicity**

Let's look at particle velocities in the 4D Minkowski spacetime (X), a measure space where gas particles move spatially and evolve longitudinally in time. Let $\mu = \mu_x$ be a measure on X, $f(\boldsymbol{x}, t) \in L^1(X, \mu)$ be an integrable function (e.g., velocity of a particle), and $T: X \to X$ be a measure-preserving transformation at position $\boldsymbol{x} \in \mathbb{R}^3$ and time $t \in \mathbb{R}^+$.

A pointwise ergodic theorem argues that in a measure theoretic sense, the average of f over all particles in the gas system at a fixed time, $\bar f = E_t(f) = \int_{\mathbb{R}^3} f(\boldsymbol{x}, t)d\mu_x$, will be equal to the average velocity (f) of just one particle ($\boldsymbol{x}$) over the entire time span,

$$\tilde f = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=0}^{n} f(T^i \boldsymbol{x})\right), \text{ i.e., (show) } \bar f \equiv \tilde f.$$

The spatial probability measure is denoted by μ_x and the transformation $T^i\boldsymbol{x}$ represents the dynamics (time evolution) of the particle starting with an initial spatial location $T^o\boldsymbol{x} = \boldsymbol{x}$.

Investigate the ergodic properties of various transformations in the 5D **spacekime**:

$$\underbrace{\bar f \equiv E_\kappa(f) = \frac{1}{\mu_x(X)}\int f\Big(\boldsymbol{x}, \underbrace{t, \phi}_{\kappa}\Big)d\mu_x}_{\text{space averaging}} \overset{?}{=} \underbrace{\lim_{t\to\infty}\left(\frac{1}{t}\sum_{i=0}^{t}\left(\int_{-\pi}^{+\pi} f(T^i\boldsymbol{x}, t, \phi)d\Phi\right)\right) \equiv \tilde f}_{\text{kime averaging}}$$

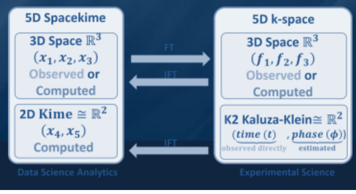
Dinov & Velev (2021)

Mathematical-Physics ⟹ Data Science & AI

Mathematical-Physics	Data/Neuro Sciences
A **particle** is a small localized object that permits observations and characterization of its physical or chemical properties	An **object** is something that exists by itself, actually or potentially, concretely or abstractly, physically or incorporeal (e.g., person, subject, etc.)
An **observable** a dynamic variable about particles that can be measured	A **feature** is a dynamic variable or an attribute about an object that can be measured
Particle **state** is an observable particle characteristic (e.g., position, momentum)	**Datum** is an observed quantitative or qualitative value, an instantiation, of a feature
Particle **system** is a collection of independent particles and observable characteristics, in a closed system	**Problem**, aka Data System, is a collection of independent objects and features, without necessarily being associated with apriori hypotheses
Wave-function	**Inference-function**
Reference-Frame **transforms** (e.g., Lorentz)	Data **transformations** (e.g., wrangling, log-transform)
State of a system is an observed measurement of all particles ~ wavefunction	**Dataset (data)** is an observed instance of a set of datum elements about the problem system, $O = \{X, Y\}$
A **particle system is computable** if (1) the entire system is logical, consistent, complete and (2) the unknown internal states of the system don't influence the computation (wavefunction, intervals, probabilities, etc.)	**Computable data object** is a very special representation of a dataset which allows direct application of computational processing, modeling, analytics, or inference based on the observed dataset
…	…

Spacekime Analytics

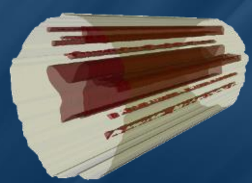
- Let's assume that we have:
  (1) Kime extension of Time, and
  (2) Parallels between wavefunctions ↔ inference functions
- Often, we can't directly observe (record) data natively in 5D spacekime.
- Yet, we can measure quite accurately the kime-magnitudes (r) as event orders, "times".
- To reconstruct the 2D spatial structure of kime, borrow tricks used by crystallographers [1] to resolve the structure of atomic particles by only observing the magnitudes of the diffraction pattern in k-space. This approach heavily relies on (1) prior information about the kime directional orientation (that may be obtained from using similar datasets and phase-aggregator analytical strategies), or (2) experimental reproducibility by repeated confirmations of the data analytic results using longitudinal datasets.



[1] Rodriguez, Ivanova, Nature 2015

Spacekime Analytics: fMRI Example

- 3D Isosurface Reconstruction of (space=2, time=1) fMRI signal



4D spacetime: Reconstruction using trivial phase-angle; kime=time=(magnitude, 0)

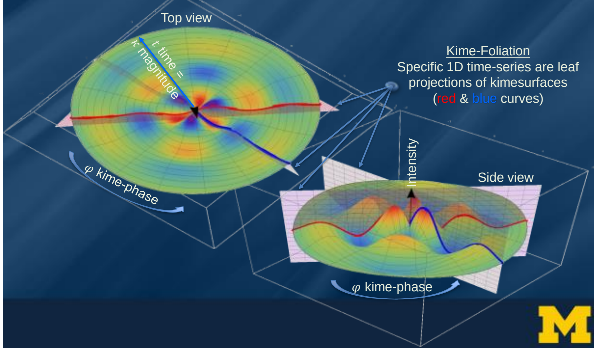
5D Spacekime: Reconstruction using correct kime=(magnitude, phase)

3D pseudo-spacetime reconstruction:

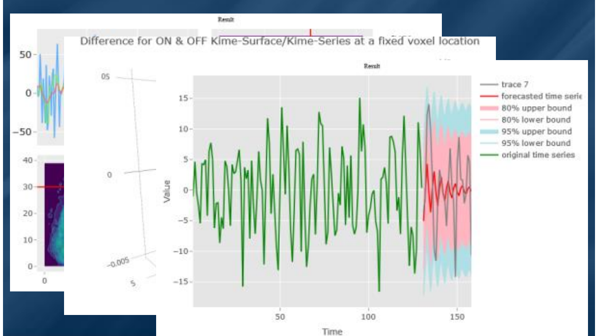
$$f = \hat{h}\,(\underbrace{x_1, x_2}_{space},\ \underbrace{t}_{time})$$

Spacekime Analytics: fMRI kime-series

fMRI *kime-series* at a single spatial voxel location (rainbow color represents fMRI kime intensities)



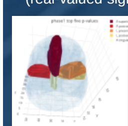
Spacetime Time-series ➔ *Spacekime Kimesurfaces*



Tensor-based Linear Modeling of fMRI

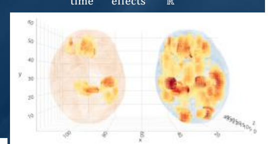
3-tier Analysis: registering the fMRI data into a brain atlas space, 56 ROIs, tensor linear modeling, post-hoc FDR processing & selection of large clusters of significant voxels are identified within the important ROIs: $Y = \underbrace{\langle X, B \rangle}_{\text{tensor product}} + E$.

The dimensions of the tensor Y are $160 \times \overbrace{a \times b \times c}^{\text{ROI b-box}}$, where the tensor elements represent the response variable $Y[t, x, y, z]$, i.e., fMRI intensity. For fMRI magnitude (real-valued signal), the design tensor X dimensions are: $\underbrace{160}_{time} \times \underbrace{4}_{effects} \times \underbrace{1}_{\mathbb{R}}$.

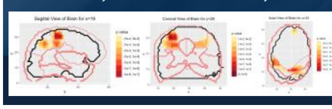


Tier 1: ROI analysis

Tier 2: Voxel analysis



Corrected Tier 3 (left) vs. Tier 2 (right): Voxel analysis



Tier 3: 2D voxel analysis projections (finger-tapping task modeling)

Spacekime Analytics: **Demos**

- Tutorials
  - https://TCIU.predictive.space
  - https://SpaceKime.org
- R Package
  - https://cran.rstudio.com/web/packages/TCIU
- GitHub
  - https://github.com/SOCR/TCIU

Acknowledgments

Slides Online: "SOCR News"

Funding

NIH: P20 NR015331, U54 EB020406, P50 NS091856, P30 DK089503, UL1TR002240, R01CA233487
NSF: 1916425, 1734853, 1636840, 1416953, 0716055, 1023115

Collaborators

- SOCR: **Milen Velev, Yueyang Shen, Daxuan Deng, Zijing Li, Yongkai Qiu, Zhe Yin, Yufei Yang, Yuxin Wang, Rongqian Zhang, Yuyao Liu, Yupeng Zhang, Yunjie Guo**
- **UMich MIDAS/MNORC/AD/PD Centers**: Chuck Burant, Kayvan Najarian, Stephen Goutman, Stephen Strobbe, Hiroko Dodge, Chris Monk, Issam El Naqa, HV Jagadish, Brian Athey

https://SOCR.umich.edu