Challenge Problem 7: Influenza-Like Illnesses

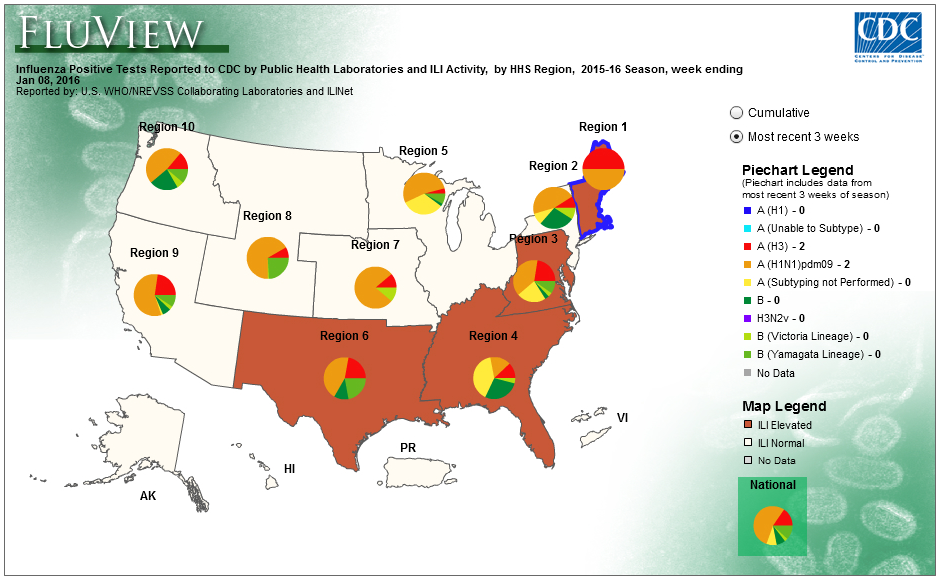
Ssu-Hsin Yu, SSCI
Tom Dietterich, Oregon State
Chad Scherrer, Galois

Influenza-Like Illnesses (ILI)

- 5 million cases of severe illness each year world wide
- 200,000-500,000 deaths annually
- Spreads by
  - contact of mucous with eyes, nose, mouth
  - inhaled aerosol particles
  - touch (e.g., hand-to-hand or hand-surface-hand)
- Virus is shed one-half to one-day after infection for a period of 5 days (longer in children and immunocompromised people)

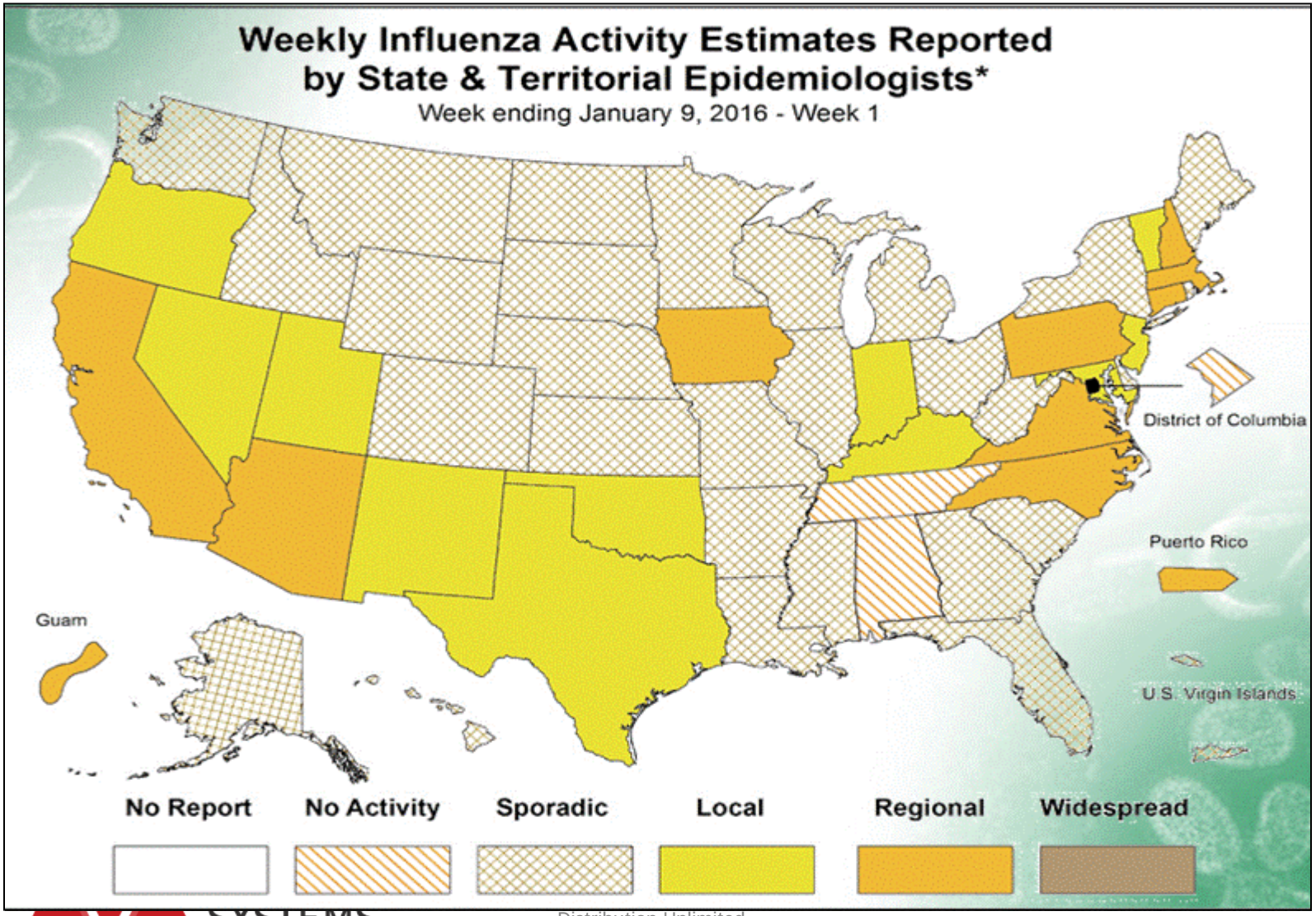
CDC U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet)

- Percentage of doctor visits that are flu-related
- Weekly reports
- Aggregated to CDC Regions
- Broken out by age ranges



http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html

Weekly State-Level Estimates



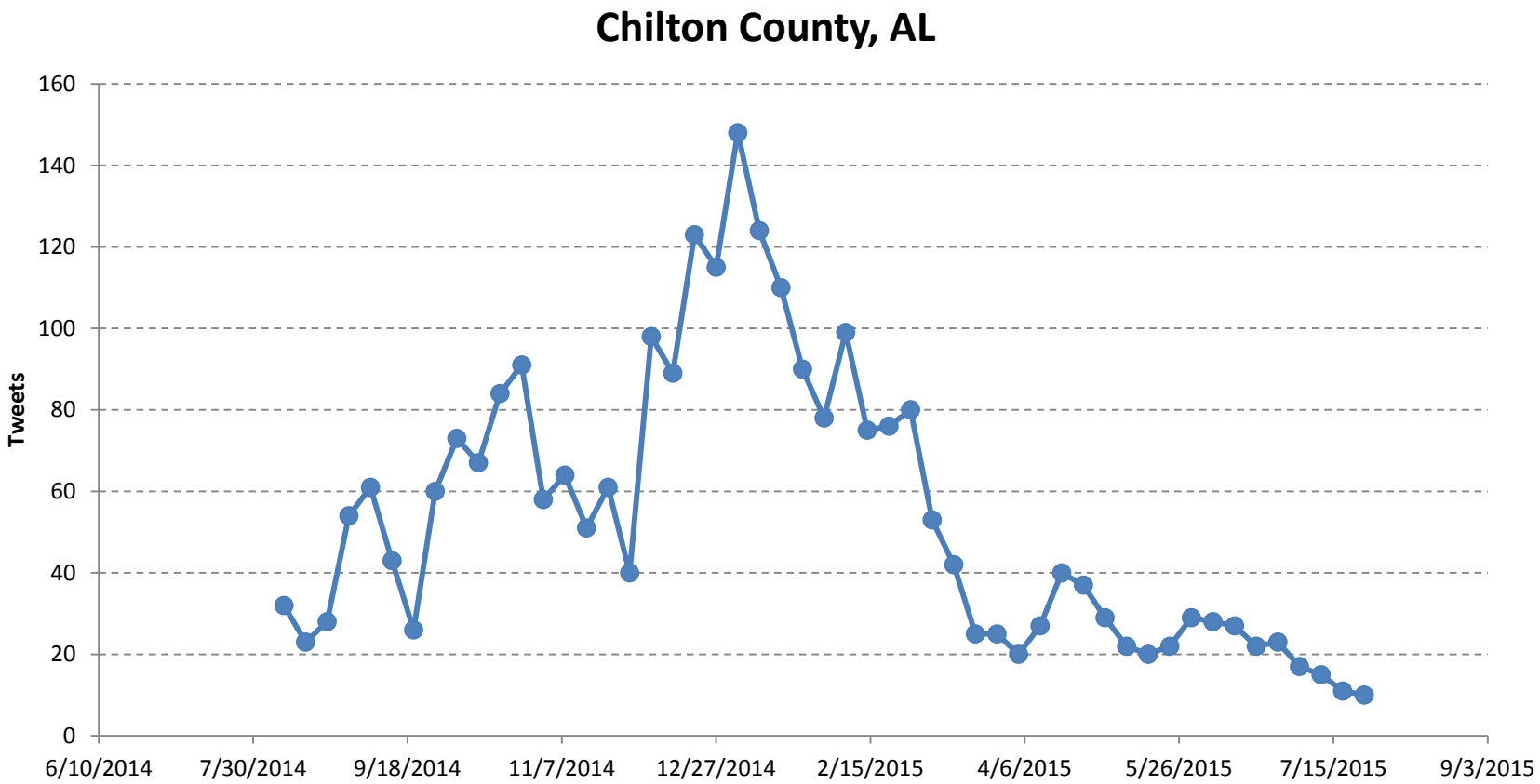
State and District Reports ("Prediction Regions")

- State-level data
  - Massachusetts, North Carolina, Rhode Island and Texas
- Within-state district data
  - Mississippi and Tennessee

Useful Covariates

- Tweets (per county per week)
  - keywords “flu” and “influenza”
  - number of tweets (not retweets)
- Cumulative Vaccination Percentage (weekly) of Medicare recipients
- Demographic information (population by age brackets)
- Geographic information: adjacent counties

Example Tweets



Proposed Model

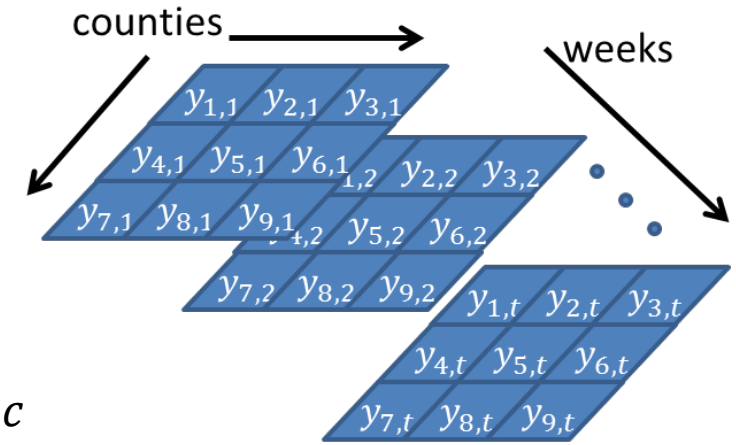
Joint multivariate Gaussian $y_{c,t}$ latent "propensity"

c: county t: week

$$p(Y) \propto \exp\left(-\tfrac{1}{2}\tau_1 Y^T (D_w - W)\, Y\right)$$

$$W = \begin{cases} w_{(c,t)(c,j)} = \rho & \text{where } j = t-1 \text{ or } t+1, \\ w_{(c,t)(i,t)} = 1 & \text{if } i \text{ is a neighboring county of } c \\ w_{(c,t)(i,j)} = 0 & \text{otherwise} \end{cases}$$

$$(D_w)_{(c,t)(c,t)} = \Sigma_{(i,j)} w_{(c,t)(i,j)}$$



Covariates

$$X_{c,t} = \left[\log\left(\frac{S_{c,t}+\epsilon_2}{\widetilde{N}_c}\right), \quad \log\left(\frac{V_{c,t}+\epsilon_3}{1-V_{c,t}+\epsilon_3}\right)\right]^{\boldsymbol{T}}$$

$S_{c,t}$: number of flu-related tweets from county *c* in week *t*.

$V_{c,t}$: cumulative percentage of Medicare recipients filing flu vaccination claims from county *c* in week *t*.

$\widetilde{N}_c = \Sigma_g N_{c,g} U_g$: Twitter user demographics adjusted population of county *c*

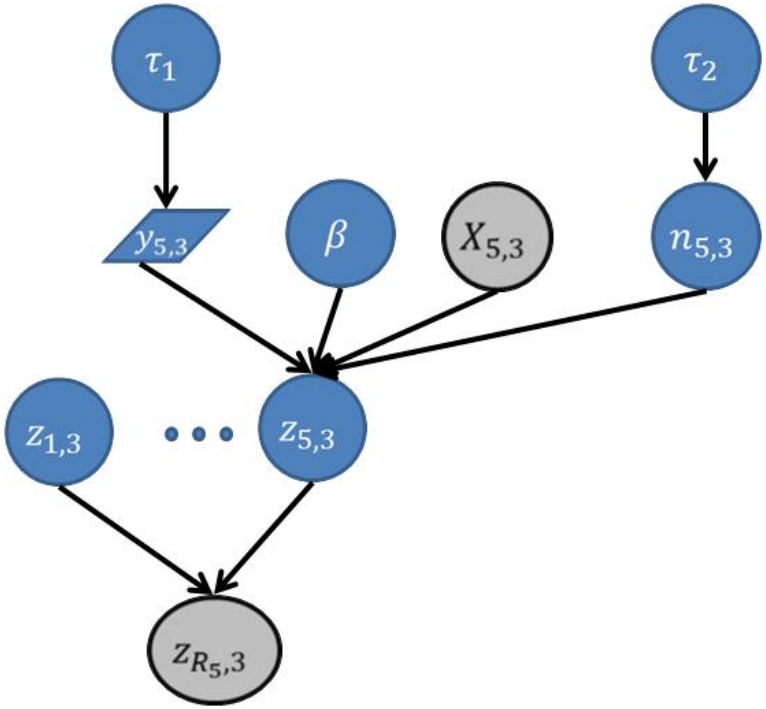
$N_{c,g}$: population of county c belonging to age group *g*.

U_g : percentage of Twitter users belonging to age group *g*.

$\epsilon_2 = 0.1$, $\epsilon_3 = 0.001$

Flu Prevalence

$$\log\left(\frac{z_{c,t}+\epsilon_1}{1-z_{c,t}+\epsilon_1}\right) = \beta^T X_{c,t} + y_{c,t} + n_{c,t}$$



c: county index; t: week index

$X_{c,t}$: covariates for each county c and week t

$y_{c,t}$: latent propensity

$z_{c,t}$: ILI rate (between 0 and 1) of county c in week t

$n_{c,t}$: zero-mean Gaussian noise with variance $1/\tau_2$

$\epsilon_1 = 0.0001$: a small number to ensure numerical stability

Aggregated Observations

$$z_{R_i,t} = \sum_{c \in R_i} \left(\frac{N_c}{N_{R_i}} \right) z_{c,t}$$

R_i : set of counties in area i; the area can be a HHS Region, a state or a district in a state

$z_{R_i,t}$: reported ILI rate of area i in week t

N_c : population of county c

N_{R_i} : population of area i

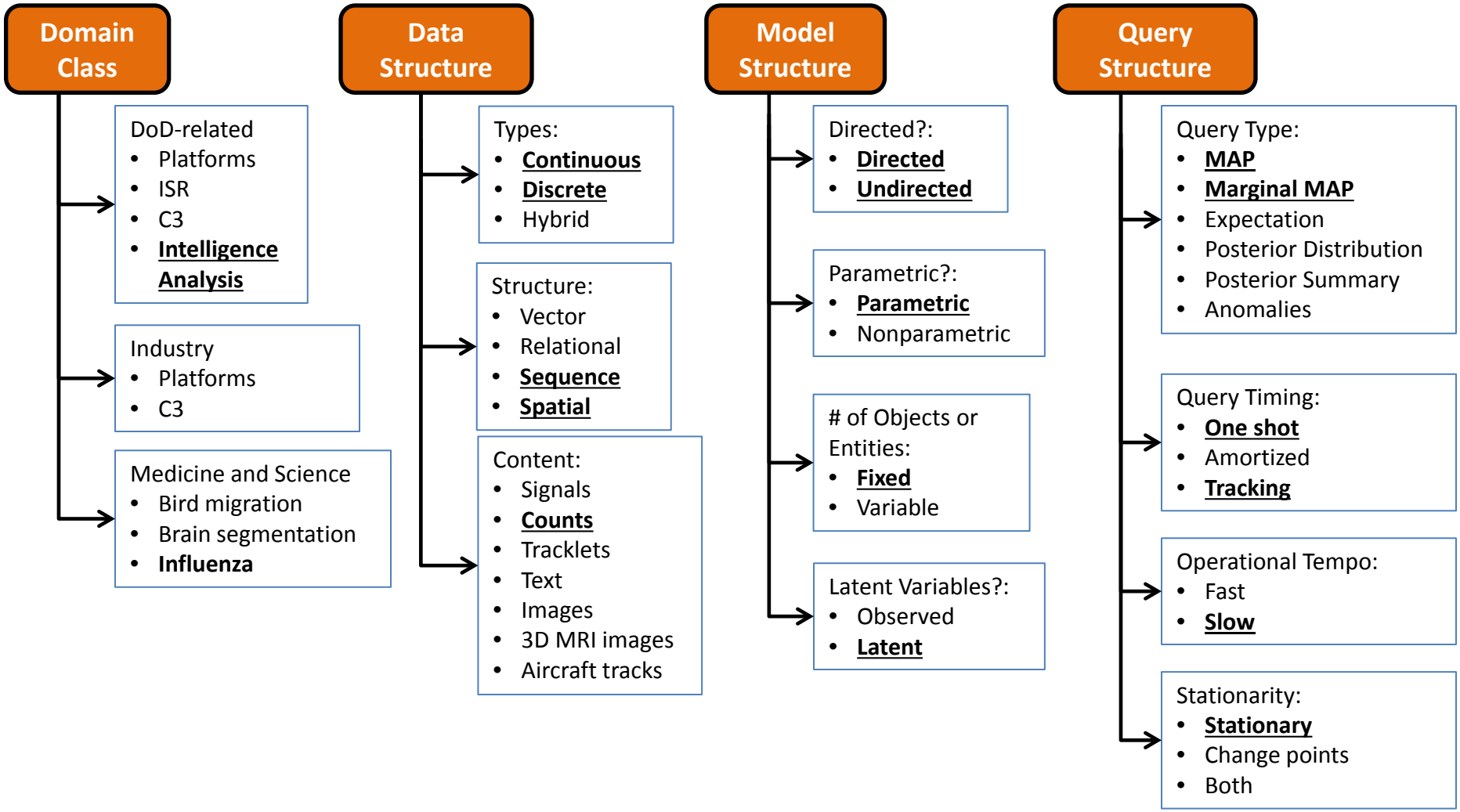
Phase 1 Task: Reconstruction

- Given:
  - weekly covariates and observations for an entire year
    - tweets, vaccination, CDC ILI reports + whole state estimates
- Find:
  - weekly ILI prevalence for all counties in the Prediction Regions
- Metrics:
  - Population-adjusted Squared Error
  - Start and Peak of the epidemic

Phase 2 Task: Weekly Nowcast for 2015-16

- Given:
  - Covariates for weeks $1, \ldots, t$
  - ILI Observations for weeks $1, \ldots, t-1$
- Find:
  - County observations for week t for the Prediction Regions
- Metrics:
  - Same as Phase 1

Challenge Problem Dimensions



CP#7 Next Evaluation Period

- Timeline
  - PI Meeting – 90 days: Beta Period Begins
  - PI Meeting – 45 days: Final Deadline for CP6 and CP7 solutions
  - July ??: PI Meeting

CP#7 Materials Available Now

- http://ppaml.galois.com/wiki/wiki/CP7FluSpread
- http://ppaml.kitware.com/midas/

Email address for questions, issues, etc.:

ppaml-support@community.galois.com

Micro-breakout ??? at ???

Future Challenge Problems

- CP8: Recognition of Interleaved Desktop Activities
- CP9: Anomaly Detection??
- CP10: Exploratory Data Analysis Hackathon??

Micro-Breakout ?? at ??

PPAML Summer School 2016

Where:

- Portland, Oregon

When:

- July 25th to August 5th, 2016

How:

- Online Announcement
  - http://ppaml.galois.com/wiki/wiki/SummerSchools/2016/Announcement
- Application Form
  - https://www.tfaforms.com/406358
- Email and forum announcements forthcoming