TRECVID 2016 Workshop
National Institute of Standards and Technology
Multimedia Event Detection Task

Nov. 15, 2016

David Joy, Jonathan Fiscus, **Andrew Delgado**

MED Session Schedule

| 9:00 – 11:20	Tuesday, Nov. 15
9:00 – 9:20	**MED Task Overview**
9:20 – 9:40	**VIREO** (City University of Hong Kong)
9:40 – 10:00	**INF** (Carnegie Mellon U.; Beijing U. of Posts and Telecommunication; U. Autonoma de Madrid; Shandong U.; Xian Jiatong U.; Singapore Management U.)
10:00 – 10:20	**MediaMill** (University of Amsterdam)
10:20 – 10:40	*Break*
10:40 – 11:00	**BUPT-MCPRL** (Beijing University of Posts and Telecommunications)
11:00 – 11:20	**MED Discussion**

Multimedia Event Detection Task

Multimedia Event Detection (MED)

Quickly find instances of events in a large collection of search videos

A MED event is a complex activity occurring at a specific place and time involving people interacting with other people and/or objects

Notional System Diagram
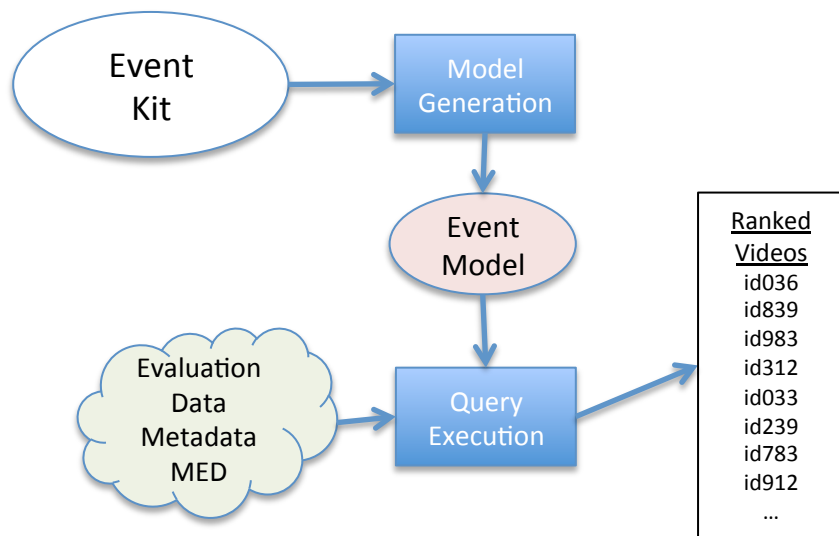
Evaluation Conditions

Execution Hardware Reporting
- 3 Classes of Computing Hardware
- Small: 100 CPU cores, 1,000 GPU cores
- Medium: 1,000 CPU cores, 10,000 GPU cores
- Large: 3,000 CPU cores, 30,000 GPU cores

Query Training Conditions

	Number of Exemplars		
Pre-Specified Events	0	10	100
Ad-Hoc Events		10	
Interactive Ad-Hoc Events		10	

Search Collection
- MED16Eval-Full -> 198K videos, 4,738 hours
- MED16Eval-Sub -> 32K video subset, 783 hours

MED '16 Overview

- MED evaluations from 2010-2016
  - Supported by the IARPA Aladdin Program and LDC collected data
  - Several simplifications in 2015, which were continued in 2016

- What's new in MED 2016
  - Introduction of new test dataset, a subset of the *Yahoo! Flickr Creative Commons 100 Million (YFCC100M) videos
  - 10 new Ad-Hoc events

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

The TRECVID MED 2016 Events

Pre-Specified Events		Ad-Hoc Events
MED '14 PS Events	MED '14 AH Events	New Events
Attempting a bike trick	Beekeeping	Camping
Cleaning an appliance	Wedding shower	Crossing a Barrier
Dog show	Non-motorized veh. repair	Opening a Package
Giving directions to a location	Fixing musical instrument	Making a Sand Sculpture
Marriage proposal	Horse riding competition	Missing a Shot on a Net
Renovating a home	Felling a tree	Operating a Remote Controlled Vehicle
Rock climbing	Parking a vehicle	Playing a Board Game
Town hall meeting	Playing fetch	Making a Snow Sculpture
Winning a race without a vehicle	Tailgating	Making a Beverage
Working on a metal crafts project	Tuning musical instrument	Cheerleading

NIST National Institute of Standards and Technology U.S. Department of Commerce

Operating a Remote Controlled Vehicle

Definition:
An individual operates a vehicle remotely with a controller

Explication:
Remote controlled vehicles are self-propelled machines that are powered by a motor or engine of some kind and whose movement is controlled from a distance by human inputs to a remote control device …

Evidential Description:

- scene: indoors or outdoors

- objects/people: remote control vehicles (cars, trucks, planes, helicopters, trains, etc.), remotes, antennas, race track, plastic takeoff ramp

- activities: directing remote control vehicles, turning on vehicles, crashing vehicles

- audio: engines revving, motor whirring, explanation of type of vehicle, discussion of where the vehicle is going or could go, discussion of what is seen on a video feed from the vehicle

Miss →

Illustrative Examples

- Positive instances of the event

- Non-Positive "miss" clips that do not contain the event

U.S. Department of Commerce
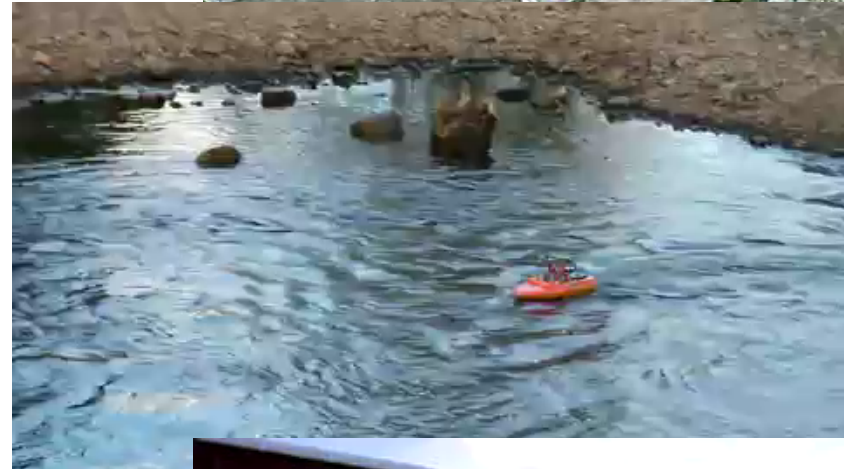
The Test Data

Data collection	Test set	# of videos	Duration (Hrs)	Avg. duration (Secs)
HAVIC Progress	MED16EvalFull	98,003	3,713	136
	MED16EvalSub	16,000	620	139
YFCC100M Subset	MED16EvalFull	100,000	1,025	37
	MED16EvalSub	16,000	163	37
Total	MED16EvalFull	198,003	4,738	86
	MED16EvalSub	32,000	783	88

- HAVIC Progress
  - Engineered target richness
  - Controlled sampling of Internet video domain
- YFCC100M Subset
  - Random selection*
  - Shorter duration videos

* - Excluding YLI-MED corpus videos (~50k videos); Excluding videos not available by mmcommons.org's AWS S3 data store (~50k videos)

12 MED 2016 Finishers By Condition

Years	Team	AH 10Ex SML	AH 10Ex MED	PS 0Ex SML	PS 0Ex MED	PS 10Ex SML	PS 10Ex MED	PS 100Ex SML	PS 100Ex MED	Organization
6	INF			Sub		Sub		Sub		Carnegie Mellon University et al.
6	MediaMill	Full				Full				MediaMill - University of Amsterdam
6	NIIHitachiUIT			Full		Sub				National Institute of Informatics
6	TokyoTech	Full				Full		Full		Tokyo Institute of Technology
5	VIREO		Full		Full		Full		Full	City University of Hong Kong & TNO
3	ITICERTH			Sub		Sub		Sub		Informatics and Telematics Inst.
3	KU-ISPL					Sub		Sub		Korea University
3	nttfudan	Full				Full		Full		NTT Media Intelligence Laboratories and Fudan University
3	MCIS					Sub		Sub		Beijing Institute of Technology Mcislab
2	BUPTMCPRL					Sub		Sub		Multimedia Communication and Pattern Recognition Labs BUPT
2	Etter			Sub		Sub				EtterSolutions
1	PKUMI							Full		Peking University
		3	1	4	1	10	1	8	1	

AH – Ad-Hoc event condition
PS – Pre-Specifed event condition

0Ex – 0 exemplar condition
10Ex – 10 exemplar condition
100Ex – 100 exemplar condition

SML – Small-sized hardware
MED – Medium-sized hardware

Full – processed MED16EvalFull test set
Sub – processed MED16EvalSub test set
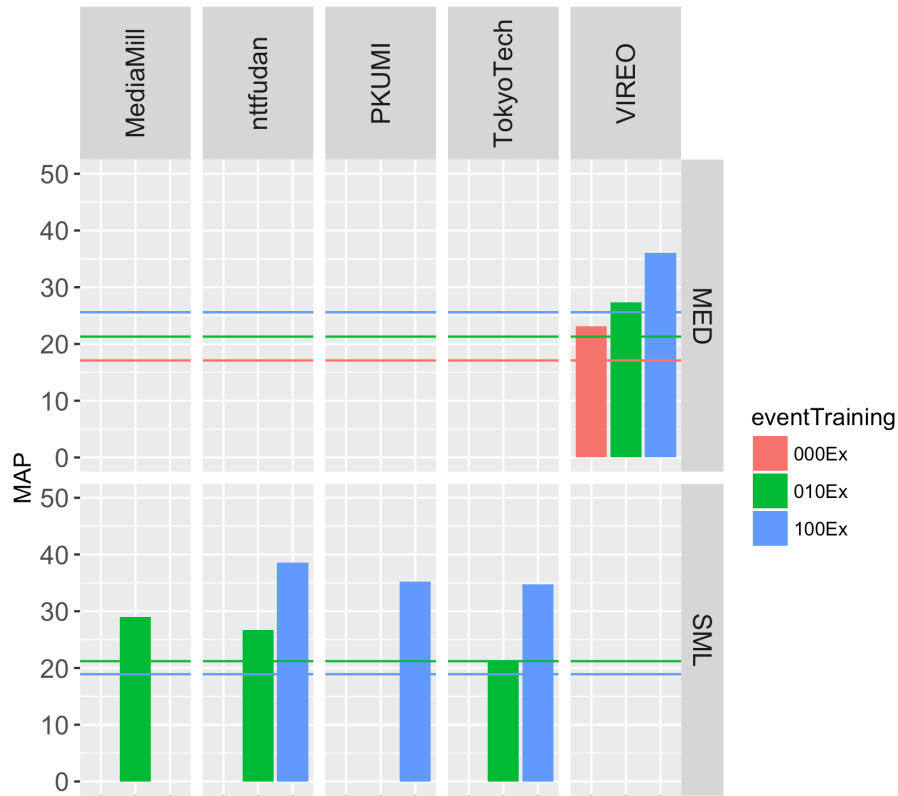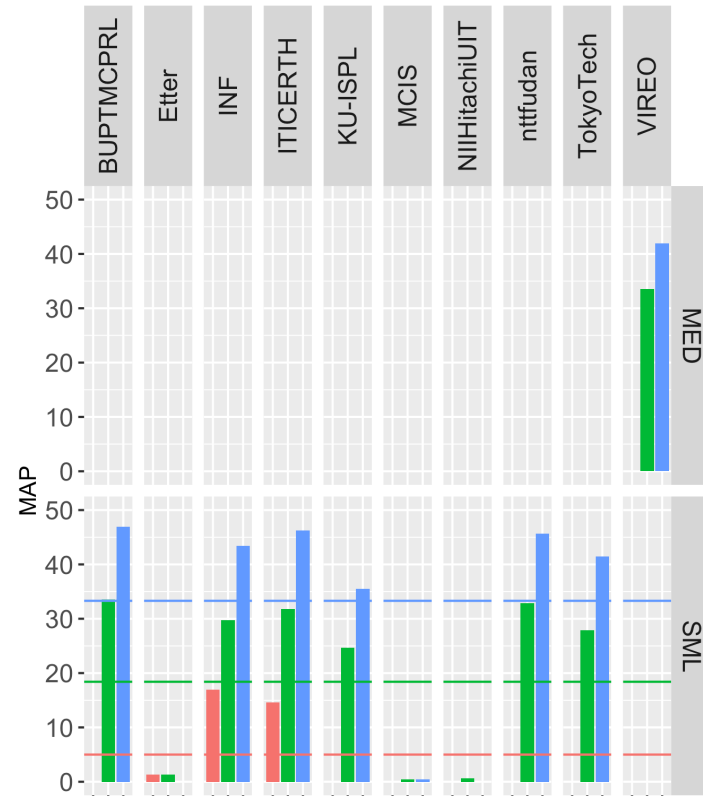- Red outline indicates a required condition

Chart: Number of MED Finishers — 2010 (Pilot), 2011, 2012, 2013, 2014, 2015, 2016 (values approximately 7, 19, 17, 18, 17, 16, 14)
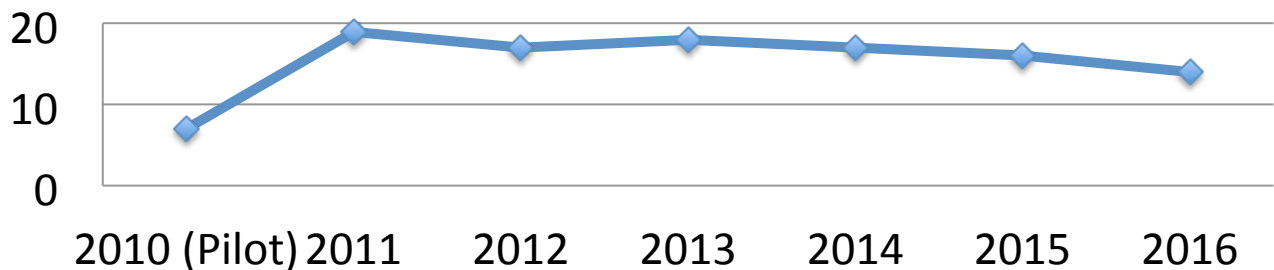
Pre-Specified Event MAP
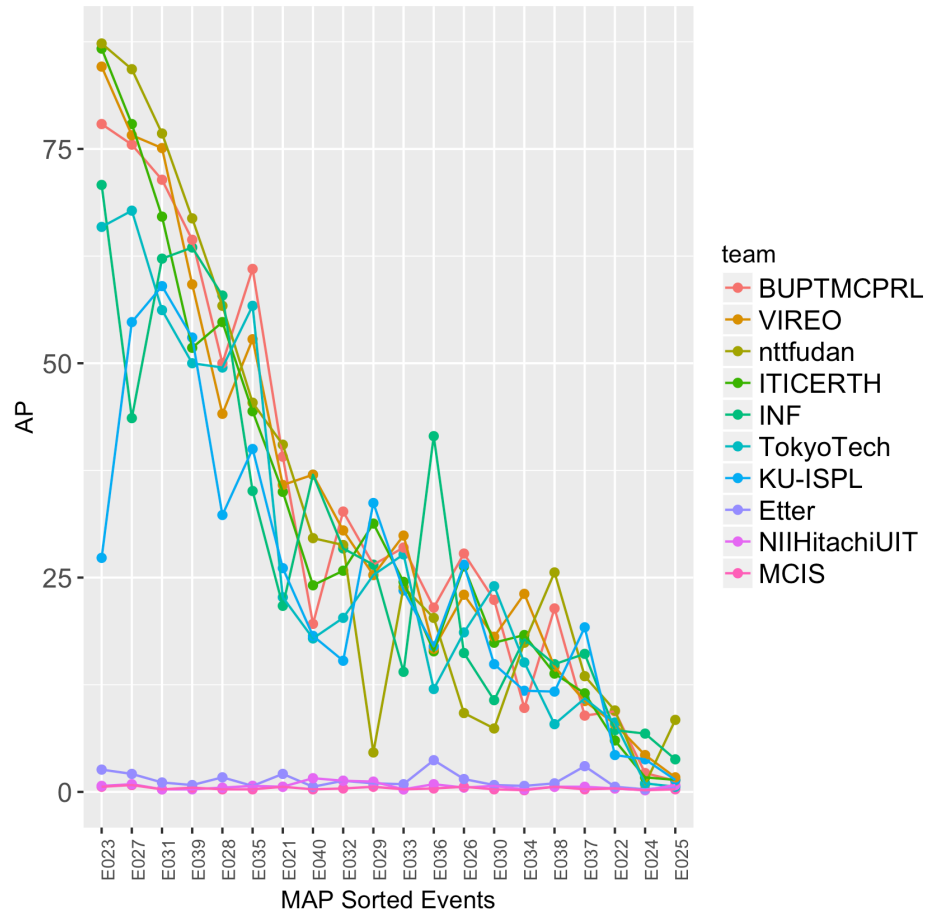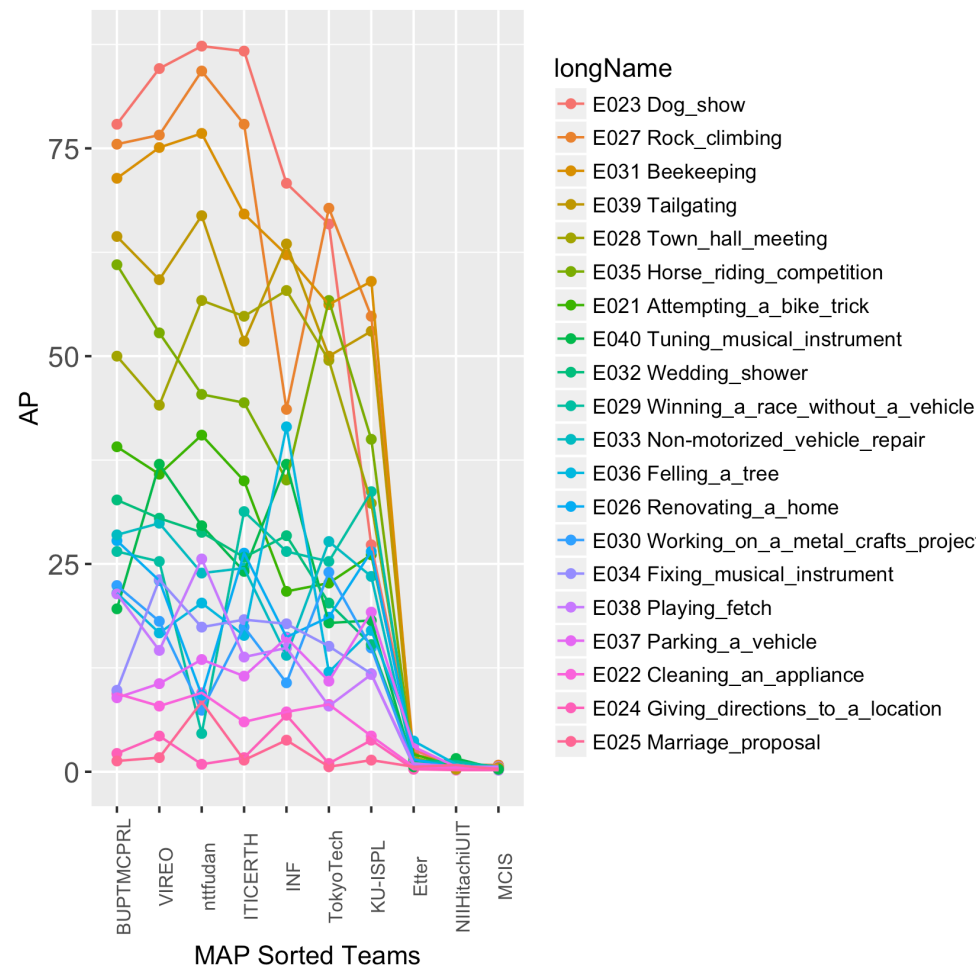## Primary Systems – HAVIC Progress Subset

EvalFull

EvalSub

MAP (EvalSub-ProgressSubset) = 1.02*MAP(EvalFull-ProgressSubset) + 5.96
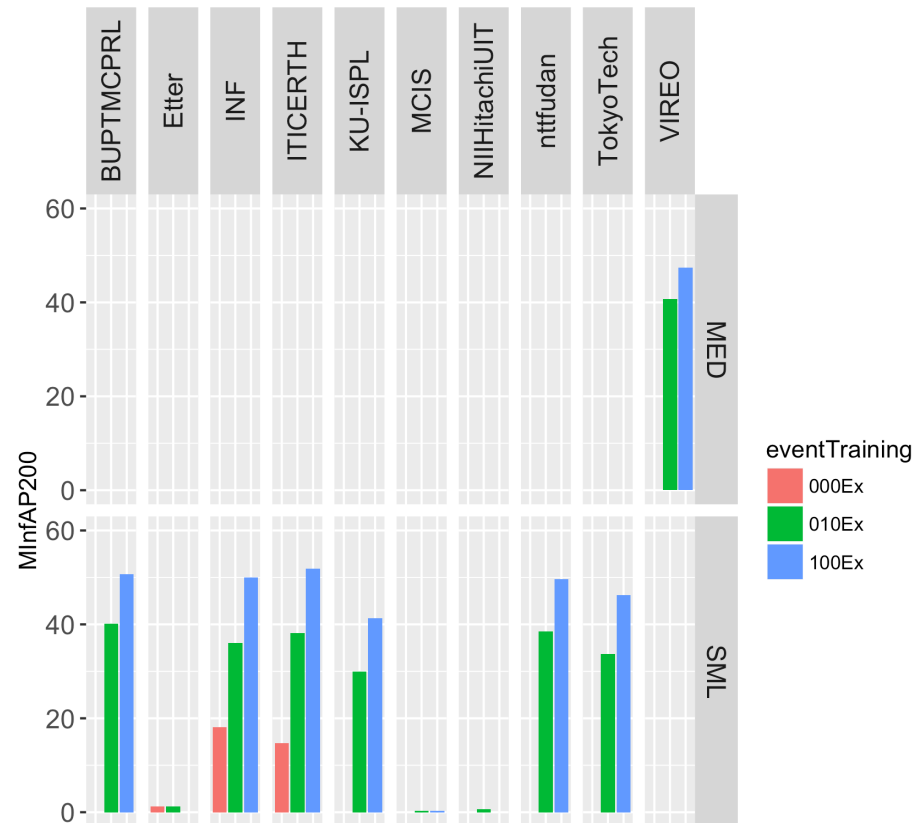R^2=0.996

Pre-Specified AP by System and Event
Primary Systems – 10Ex – MED16EvalSub – Mixed System Size
Progress Subset

MAP → Mean Inferred Average Precision (MInfAP)

- Follows Aslam et al. procedure, Statistical Method for System Evaluation Using Incomplete Judgments Proceedings of the 29th ACM SIGIR Conference, Seattle, 2006.
  - Stratified, variable density, pooled assessment procedure to approximate MAP
- MInfAP in the 2016 evaluation
  - Progress – MAP and MInfAP200 (simulated) on PS and AH
  - Progress + YFCC100M – MInfAP200 on PS and AH
- For MED '15, NIST ran experiments with 2014 data to optimize the strata sizes and sampling rate. This same sampling rate was used for MED '16
  - Define 2 strata
    - 1-60 -> 100 %
    - 61-200 -> 20 %

Pre-Specified EvalSub Simulated MInfAP200 Progress Subset

PS-EvalSub-ProgressSubset -- MInfAP200 (Simulated) = 1.14*MAP + 0.421
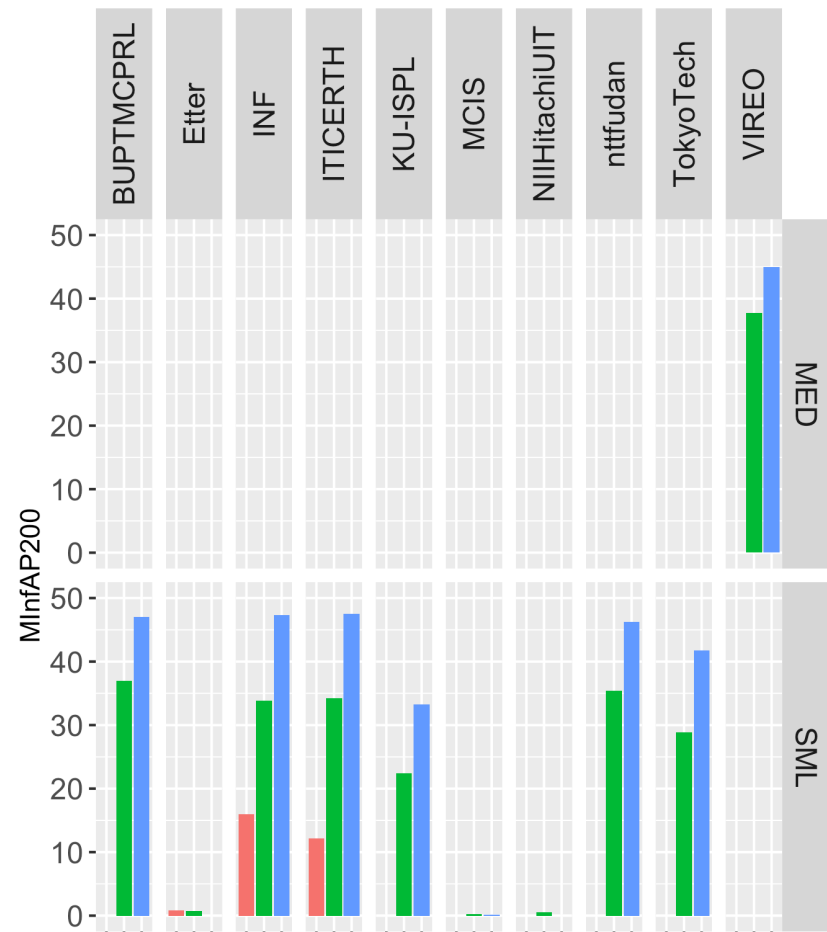$R^2=0.99$

Pre-Specified Event MInfAP200
Progress + YFCC100M
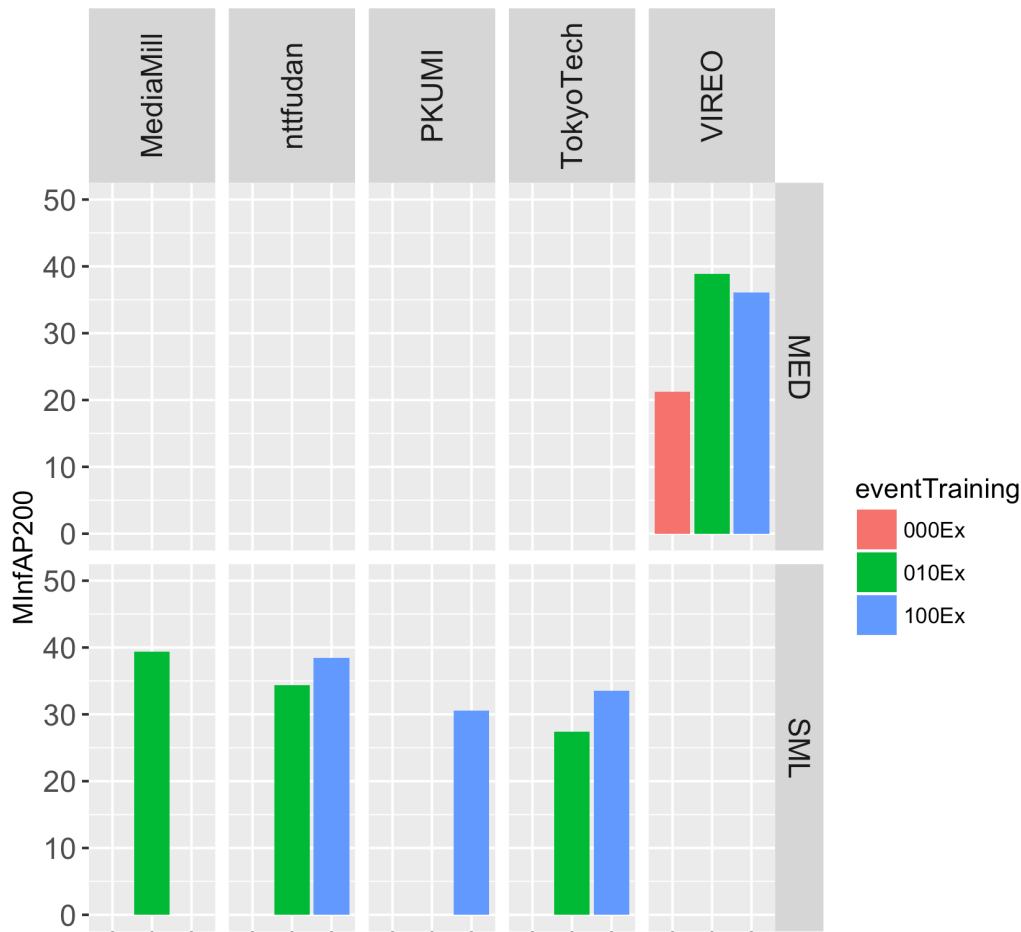
Performance on HAVIC vs. Yahoo!
Our Attempt to Score Precision@10

- Unable to score Precision @ 10 for both HAVIC and Yahoo!
  - Stratified sampling did not yield sufficient judgements
- For example:
  - E022 – Cleaning an appliance
  - Proportion of subset of top 200 clips, binned by rank
  - Teams shown are representative based on MInfAP200 scores
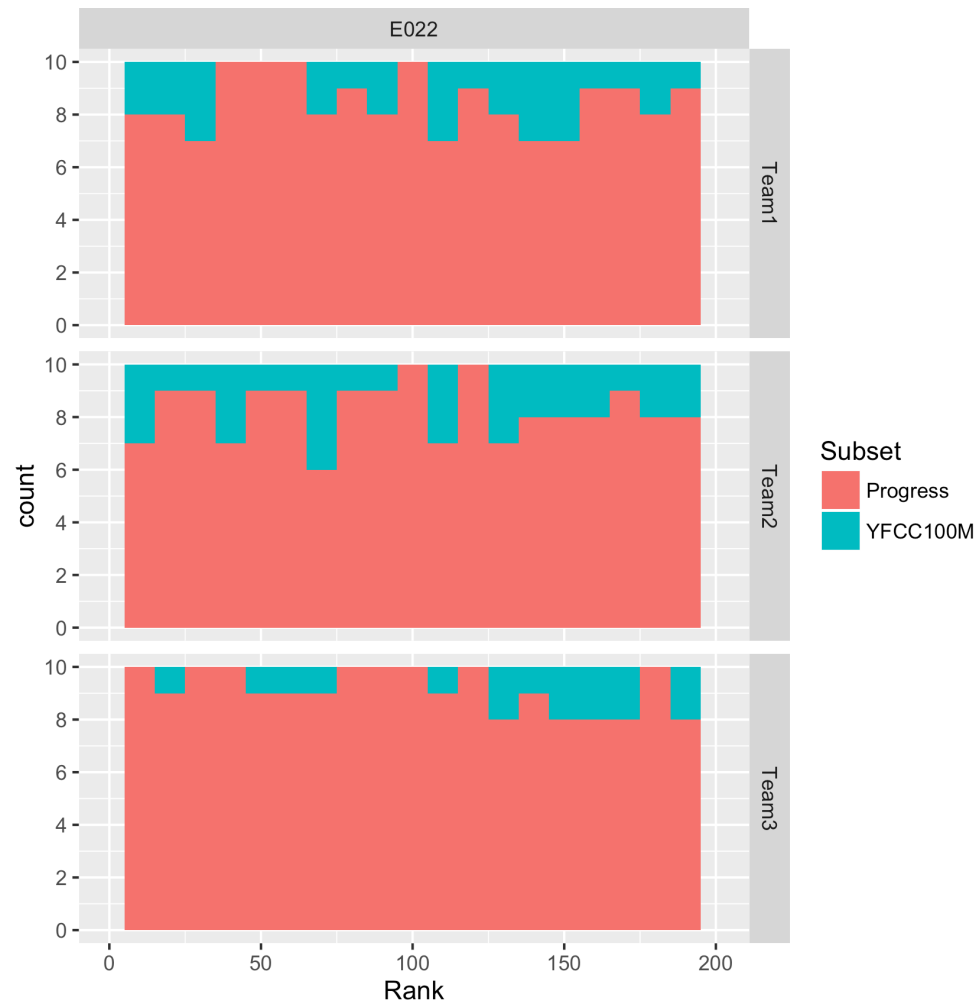
Performance on HAVIC vs. Yahoo!
Good, Bad, Ugly Events
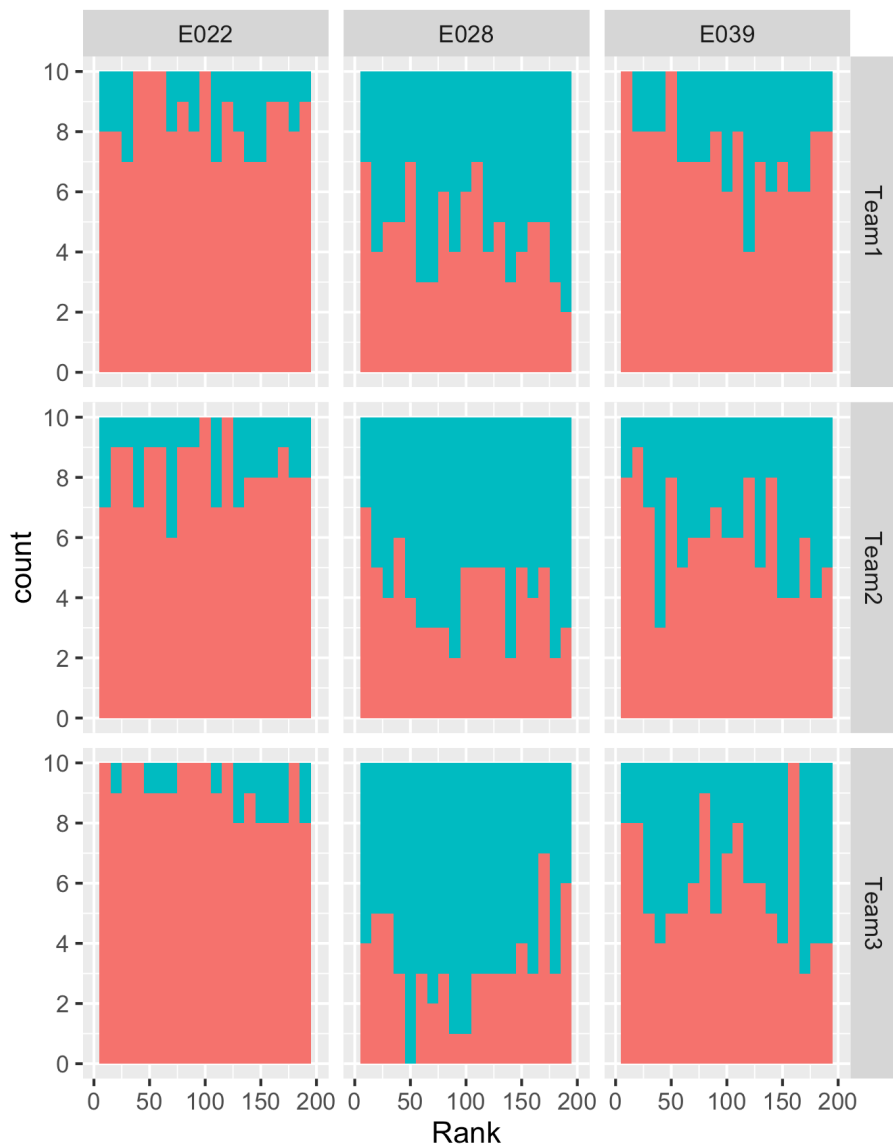
| Event	AP(top5)	Description
E022	8.42	Cleaning an appliance
E028	53.78	Town hall meeting
E039	61.4	Tailgating

- Stratified random sample not sufficient for heterogeneous data
- We will continue to work on scoring Yahoo! separately

Ad-Hoc Event Results

- 10 new events
  - 10 Exemplar training only

- MED16EvalFull the required condition

- Reference Generation
  - Pooled assessment with using all submissions
  - Strata definition
    - 1:60:100%
    - 61:200:20%

Ad-Hoc MInfAP200 on EvalFull
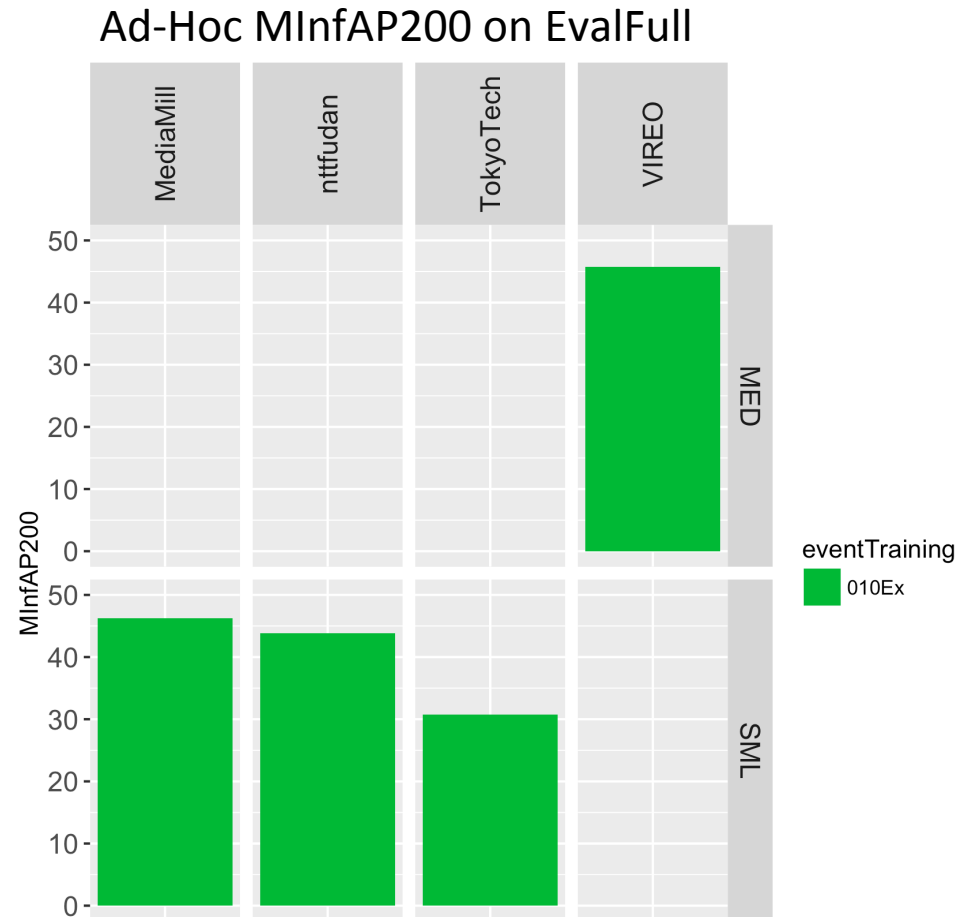
Ad-Hoc InfAP by System and Event
Primary Systems – 10Ex – MED16EvalFull – Mixed System Size
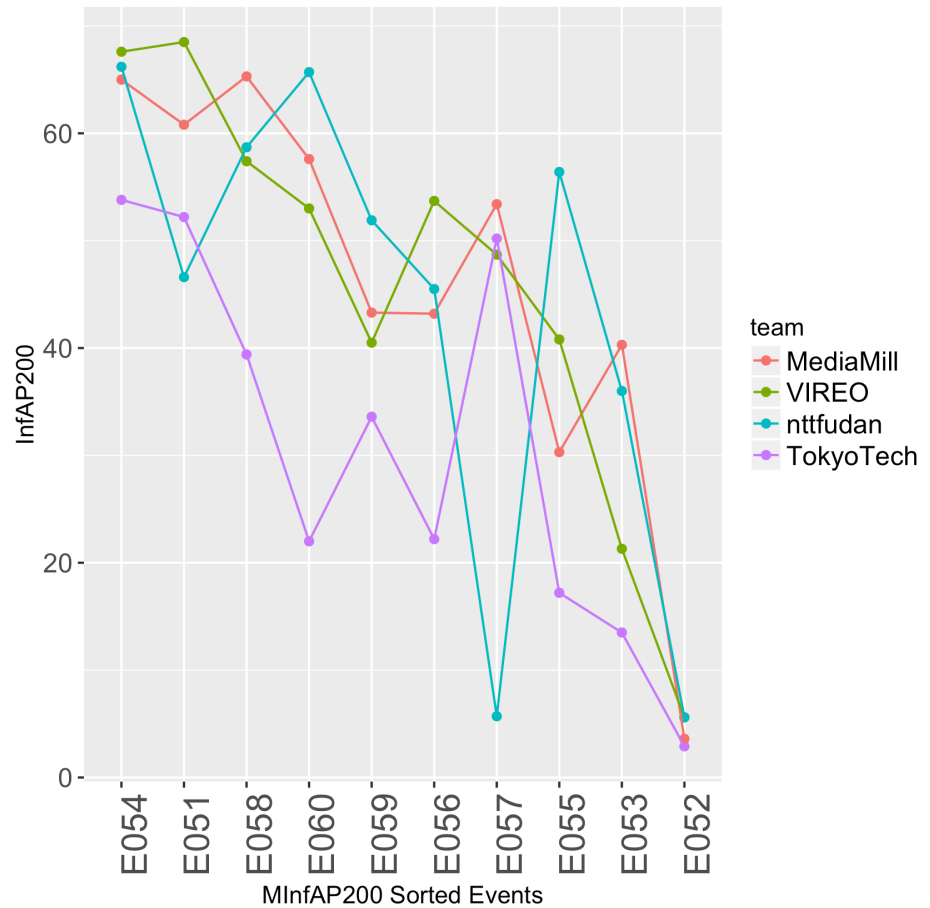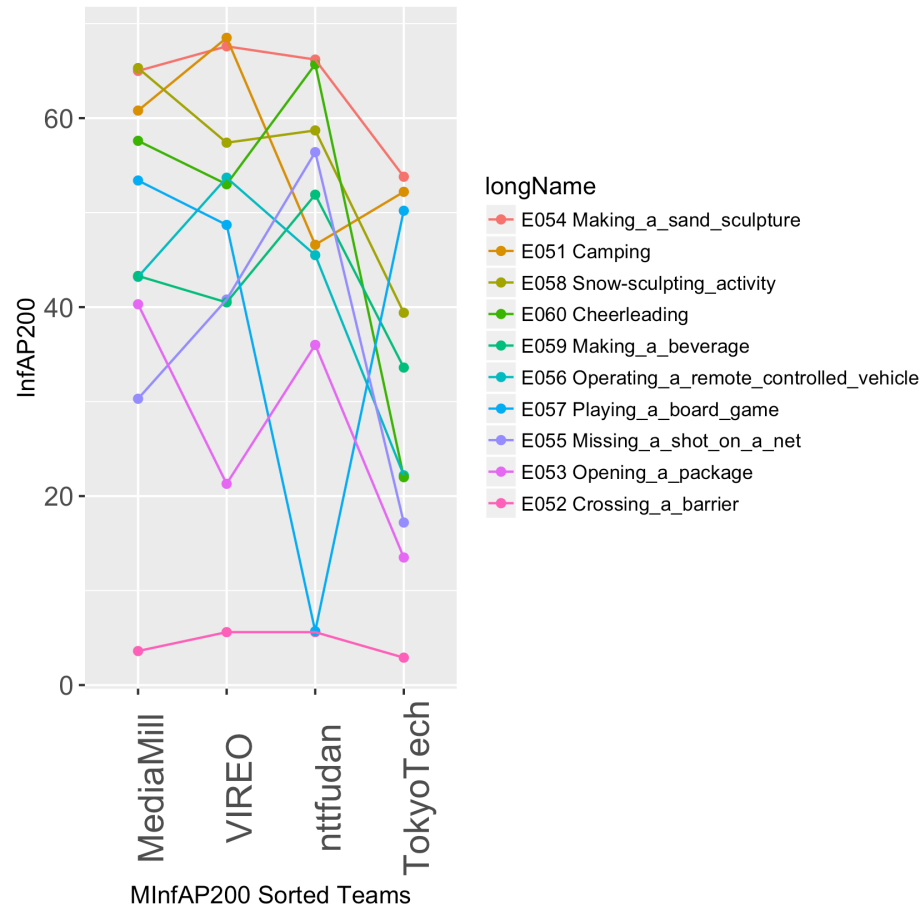
Ad-Hoc Pooled Assessment Event Richness vs. InfAP

E051	Camping
E052	Crossing a Barrier
E053	Opening a Package
E054	Making a Sand Sculpture
E055	Missing a Shot on a Net
E056	Operating a Remote Controlled Vehicle
E057	Playing a Board Game
E058	Making a Snow Sculpture
E059	Making a Beverage
E060	Cheerleading

Event Richness in Annotation Pools
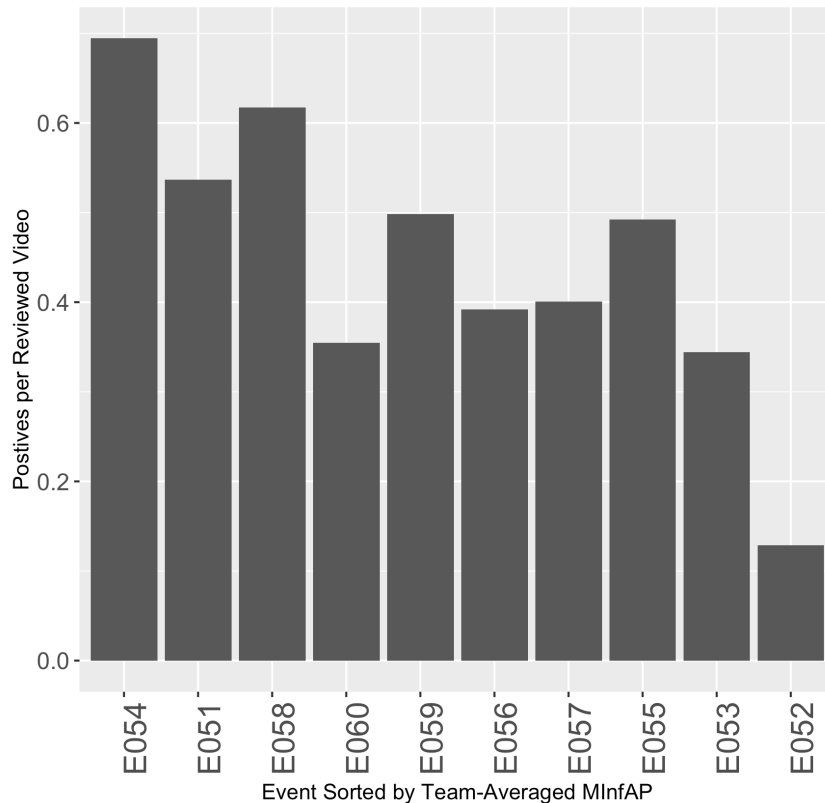
Ad-Hoc 10Ex – MInfAP200 Box Plots

National Institute of Standards and Technology
U.S. Department of Commerce
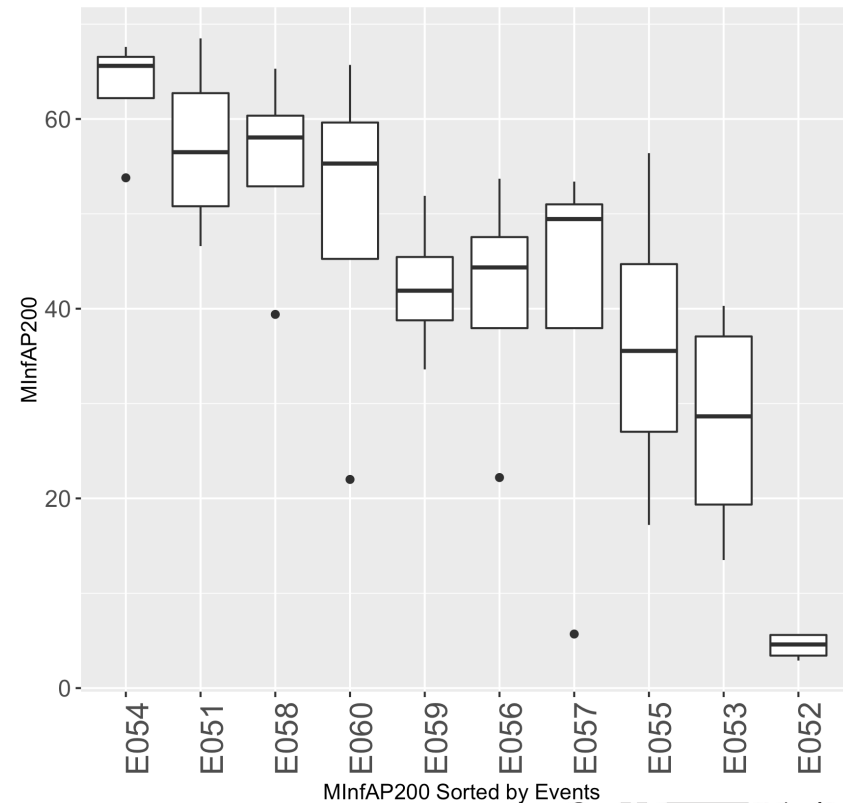
MED '16 Summary

- Pre-Specified Results
  - Only one team built a "Medium" hardware system
  - Most teams processed the subset (783 hr.) test set (MED16EvalSub)
  - Noticeable improvement over last year's Pre-Specified results on Progress
  - Stratified random sampling on heterogeneous data not powerful enough to determine differences in performance
- Ad-Hoc Results
  - Only 4 of 12 teams participated
  - No teams participated in Interactive Event Query test

MED '17 Plans

- NIST intends to continue MED in a streamlined fashion
- NIST to release Progress annotations
- Makeup of data sets TBD, HAVIC + YFCC100M
- Discontinuing support for Interactive Ad-Hoc condition
- Counter-proposals?

Thank you!

Questions?