

# Identification of tourist flows in Florida to support development of tourist travel module for FDOT Florida Transportation Model

PM: Thomas Hill

PI: Andrei Kirilenko

Co-PI: Jin Won Kim

PhD students: David Ma, Eunjung Yang

University of Florida

Contract number: BDV31 TWO 977-118 / AWD06352 (UF)

# Outline

- Project overview
- Tourism trips in transportation model
- Tourism related infrastructure in Florida
- Social media data
- Cell phone data
- Conclusion

# Background

- Florida is the largest tourism destination worldwide, receiving 125 million visitors in 2018
  - Domestic: 110M; foreign: 14M
- Tourism travel to and within Florida has substantial impact on the state's highway system
- These flows currently are not fully incorporated into FLSWM
  - location and capacity of all major tourism resources (including tourism attractions, accommodations, etc.)
  - new sources of data: social media, mobile phone tracking
- Overarching problem: identification of traffic related to tourism
  - Anecdotal evidence that understanding road traffic related to tourism is important for management
  - Existing data is patchy; can we fill in the gaps using secondary data (social networks, mobile devices) and modeling?

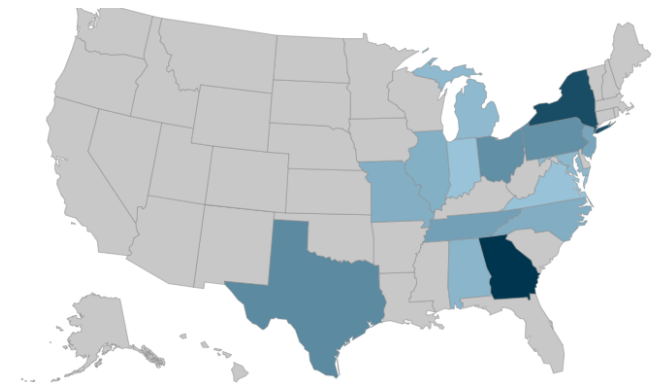
# Introduction

- Overarching problem: identification of traffic related to tourism
  - Anecdotal evidence that understanding road traffic related to tourism is important for management
  - Existing data is patchy; can we fill in the gaps using secondary data (social networks, mobile devices) and modeling?
- Department of Tourism, Hospitality and Event management and the University of Florida
  - Tourism Analytics
  - Introduction of UF participants



# Background

- Florida is the largest tourism destination worldwide, receiving 125 million visitors in 2018
  - Domestic: 110M; foreign: 14M
- Tourism travel to and within Florida has substantial impact on the state's highway system
- These flows currently are not fully incorporated into FLSWM
  - location and capacity of all major tourism resources (including tourism attractions, accommodations, etc.)
  - new sources of data: social media, mobile phone tracking
- Problem: methodology is not developed
  - Need research focusing on integrating tourism travel into a statewide transportation model

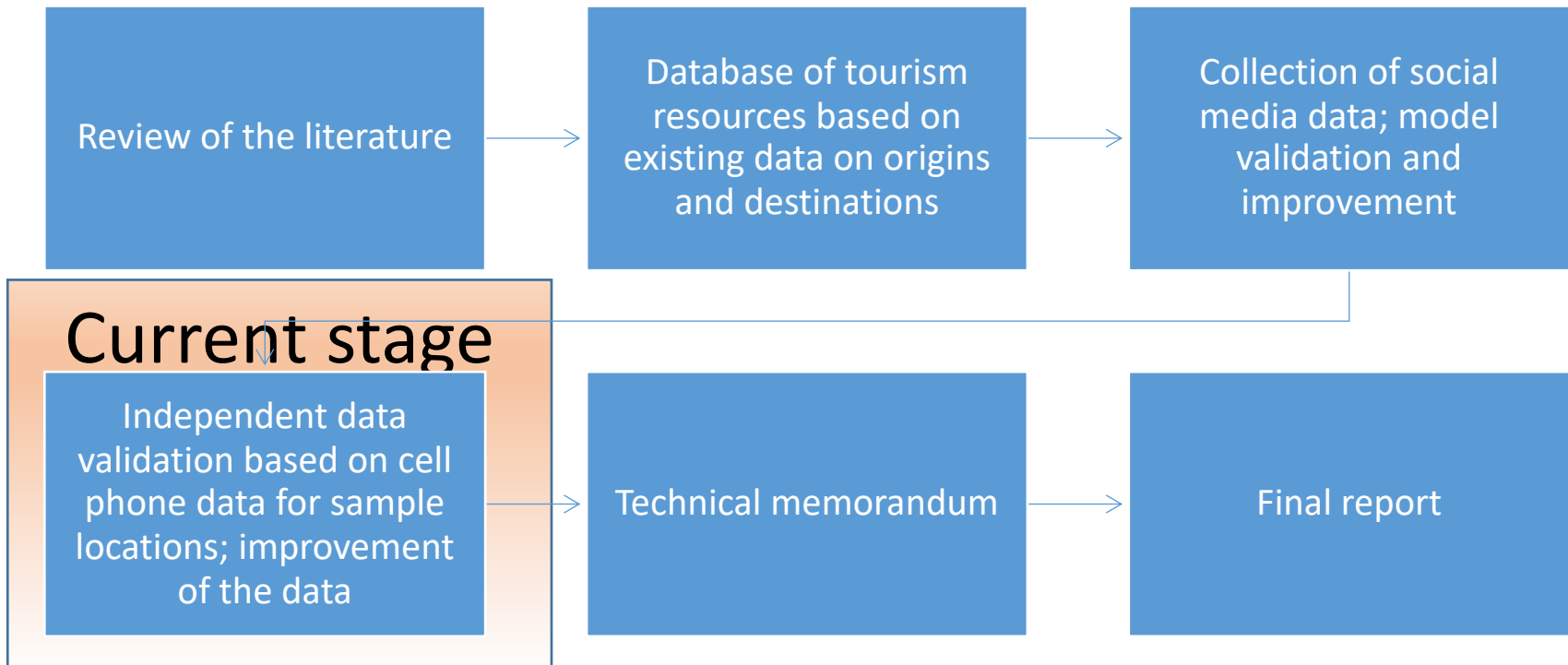


Top origin states

# Project overview

- Goal: to develop data and tools needed to integrate **tourism related** travel into transportation model
- Objectives
  - Present methodology for incorporation tourism travel into the Florida travel model
  - Collect data contributing to understanding tourism flows in Florida
  - Evaluate and cross validate tourism related data and build origin-destination trip matrices to incorporate into the trip model

# Overall project structure



# Tourist observation data

<b>data source</b>	<b>advantages</b>	<b>disadvantages</b>
surveys	detailed info on traveler richness of collected data	time-consuming, costly small sample very poor resolution
GPS trace	extremely fine spatial and temporal resolution	need to recruit in advance no data on traveler unknown representativeness
mobile phone networks	large volume fine spatial and temporal resolution	potential invasion of privacy no info on individual traveler costly
geotagged social media	low privacy concerns low cost, large volume	potential sample bias
travel journals	less privacy issues rich data on visitations	non-representative sample wide range in data quality

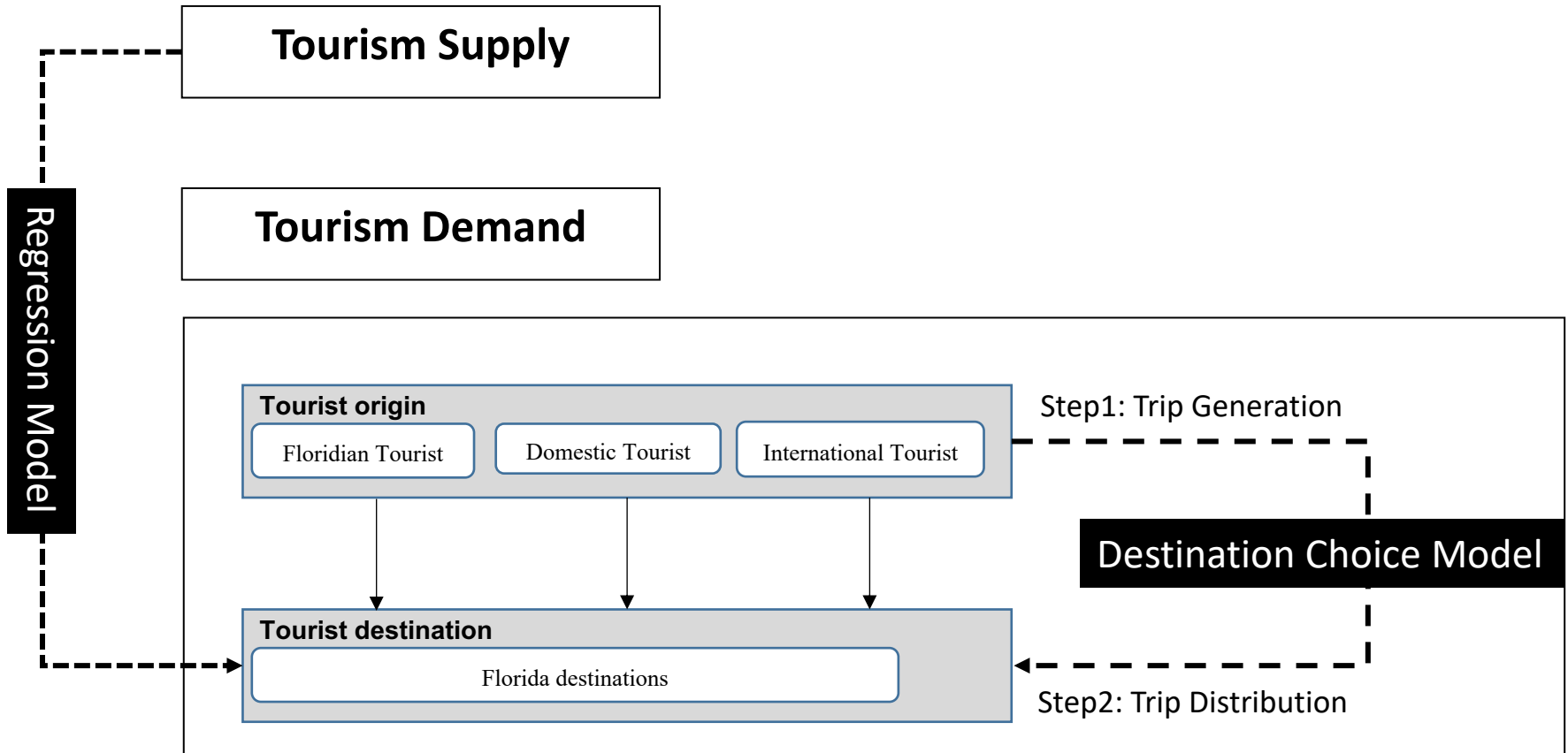
# Tourist observation data

data source	advantages	disadvantages
surveys	detailed info on traveler richness of collected data	time-consuming, costly small sample very poor resolution
GPS trace	extremely fine spatial and temporal resolution	need to recruit in advance no data on traveler unknown representativeness
mobile phone networks	large volume fine spatial and temporal resolution	potential invasion of privacy no info on individual traveler costly
geotagged social media	low privacy concerns low cost, large volume	potential sample bias
travel journals	less privacy issues rich data on visitations	non-representative sample wide range in data quality

# Data sources in this project

- Fine spatial resolution social media data from TripAdvisor reviews of Florida attractions posted by tourists countrywide
  - Advantage: we know they are tourists; we can easily collect new data to follow dynamics
  - Disadvantage: unknown biases in data; academic literature still not settle on correction of biases
  - Usability: Initial estimation tourist travel from different locations and between attraction points;
- Locational intelligence data
  - Advantage: presumably no bias (for domestic travelers)
  - Disadvantage: no internal validation mechanism for tourist travel
  - Usability: validation and enhancement of the social media data;
- Survey data
  - Advantage: direct observation; actual data on travelers
  - Usability: validation and calibration of indirect observational data.

# Overall Methodology Framework



# Tourism Supply/Demand Modeling approach

## Tourism Supply

Tourism Resources

- Tourism resource indices
- Tourism attractions (social media)

Tourism Industry

- Lodging (Hotel & Airbnb)
- Food/beverage (Restaurant)

Transportation System

- Highway accessibility/availability

## Tourism Demand

Tourist visitation

- Mobile phone data
- Social media data



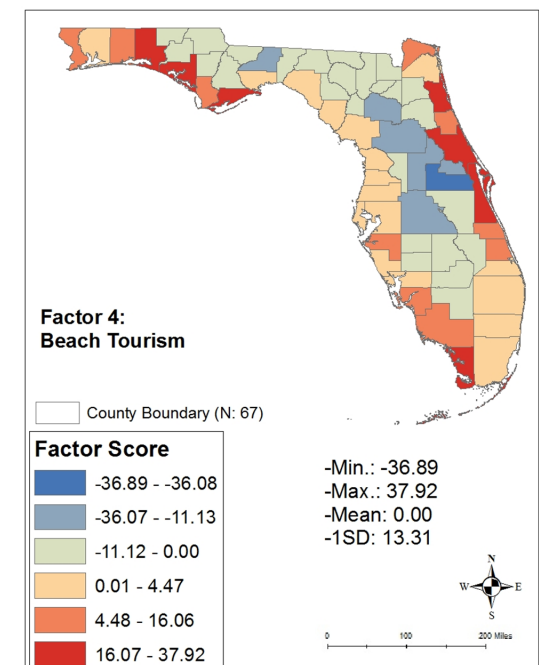
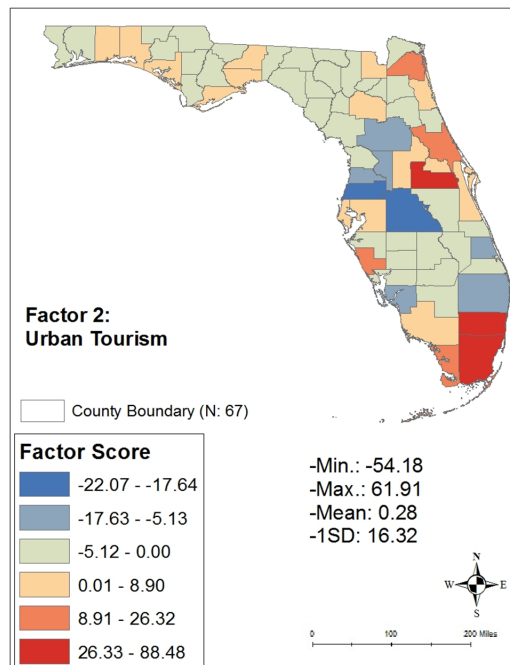
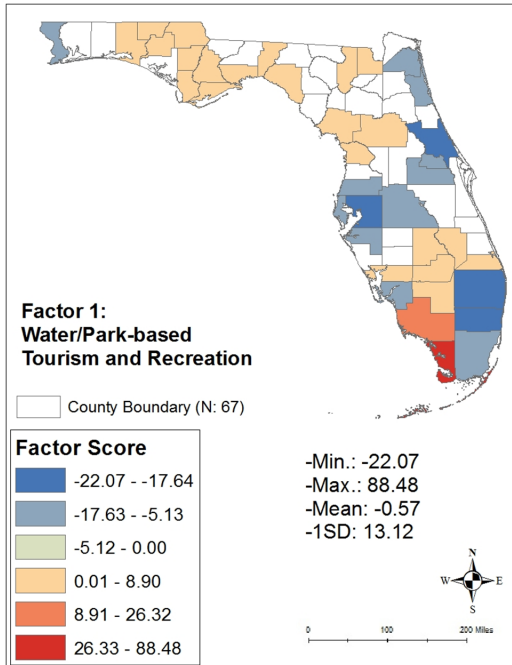
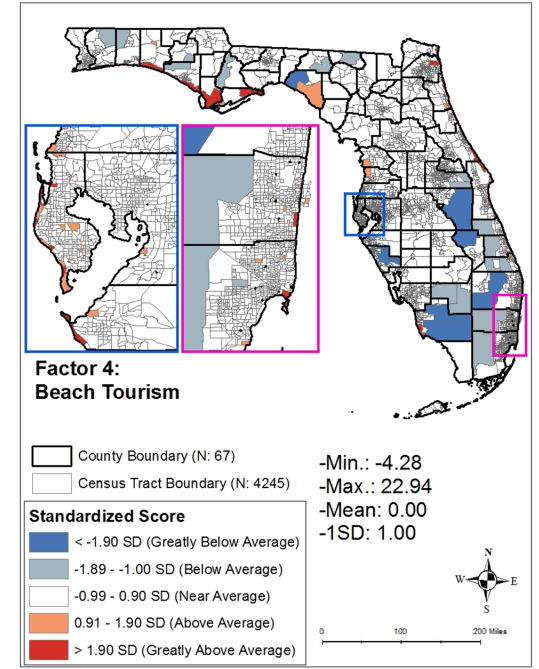
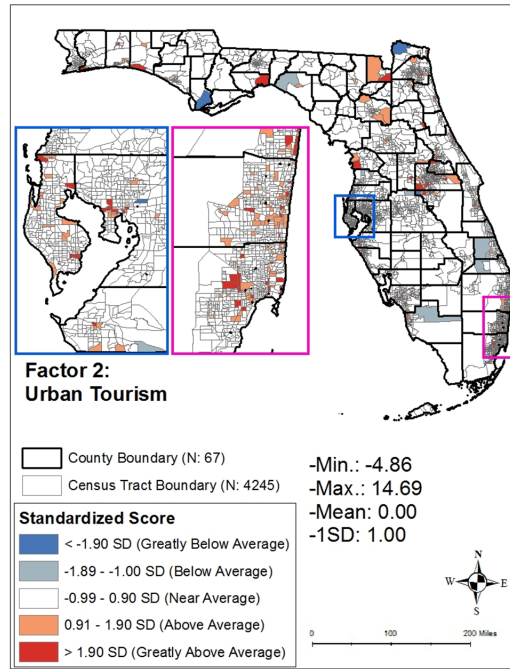
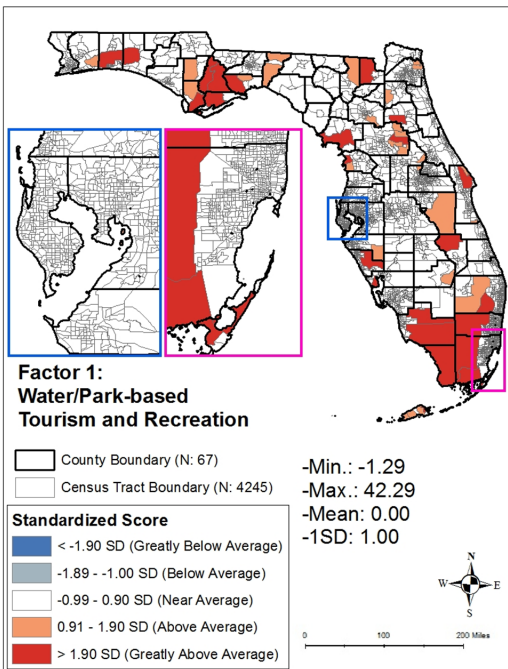
Geospatial Approach

- Spatial correlation
- Spatial regression



# What factors determine Florida tourism supply? Tourism Resource Indices

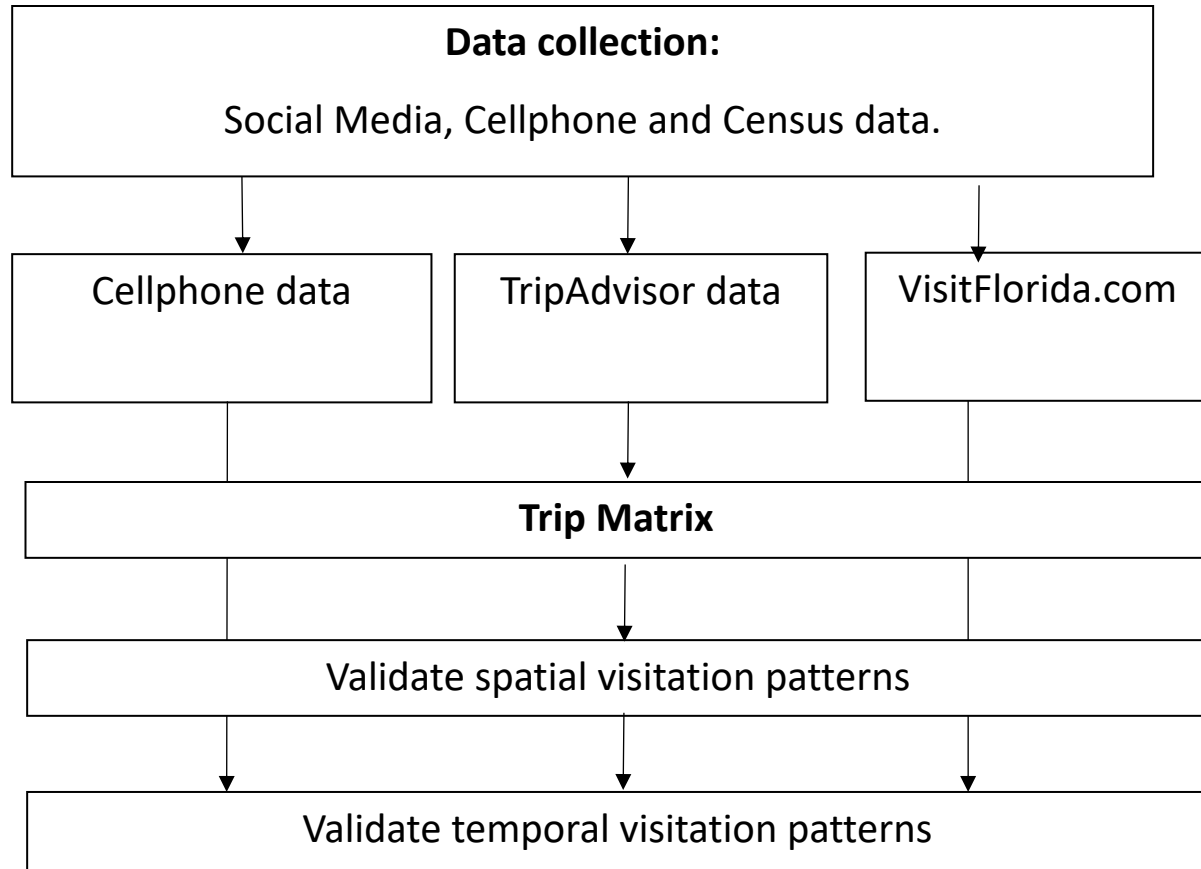
- FDOT geo-coded tourism inventory data: 92 fields including tourist origins and destinations
- 12 tourism resource factors identified (census tract & county resolution)
  - Factor 1: Water/Park-based Tourism and Recreation
  - Factor 2: Urban Tourism
  - Factor 3: Recreational Boating Tourism
  - Factor 4: Beach Tourism
  - Factor 5: Golf Tourism
  - Factor 6: RV and Camping
  - Factor 7: Horse/Race Track
  - Factor 8: Aquarium/Zoo Tourism
  - Factor 9: Theme Park/Casino Tourism
  - Factor 10: Sport Tourism
  - Factor 11: Garden Tourism
  - Factor 12: MiCE Tourism



# Four-step modelling approach

- Traditional four-step modelling approach:
  - **trip generation**
  - **trip distribution**
  - model choice
  - route assignment
- Data requirements for destination choice model:
  - origin census feature
  - destination attractiveness
  - origin – destination flows (OD matrix)
- Tourist Segments: international, domestic, Floridian

# Data Processing For DCM

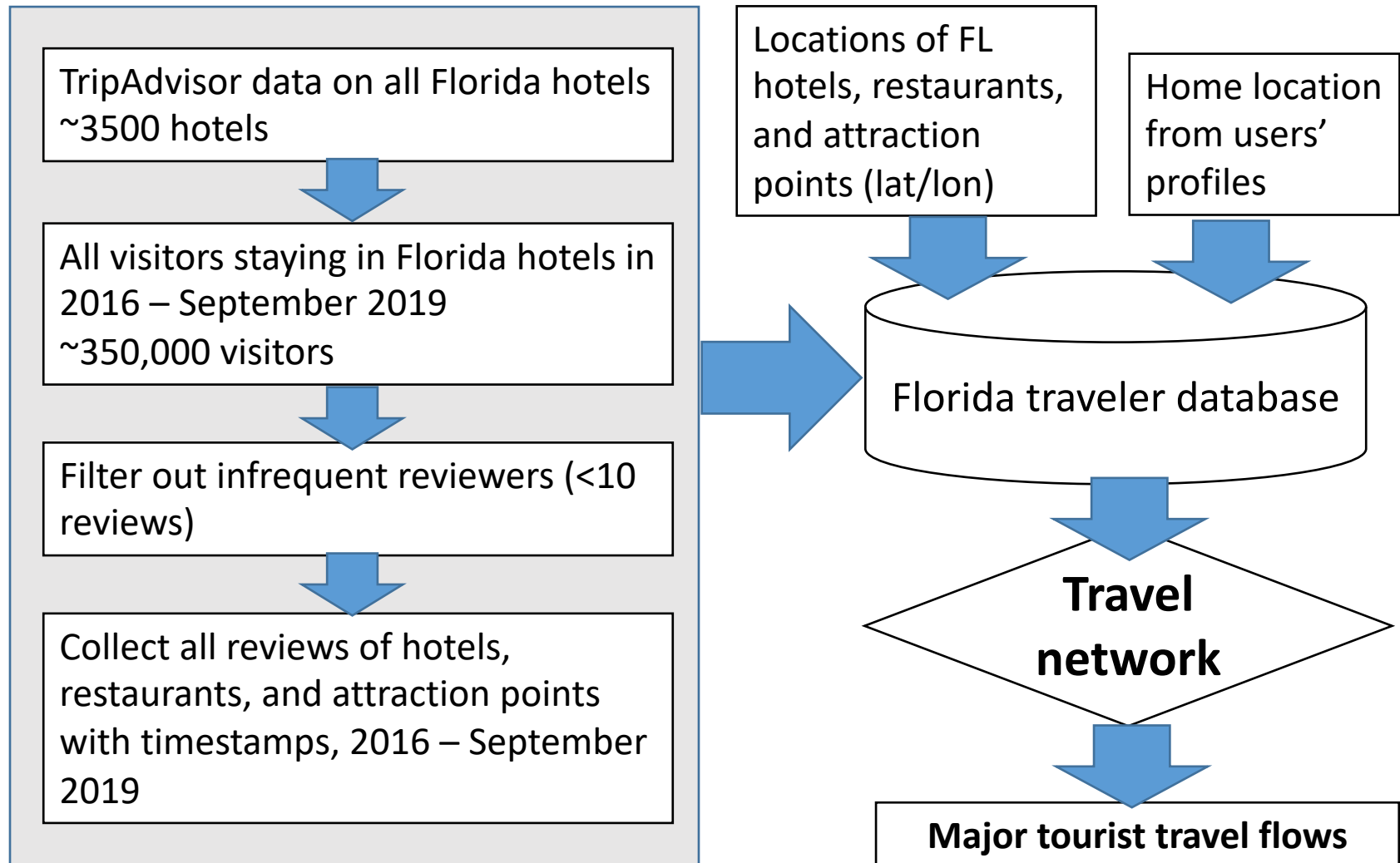


# Traditional visitation data (VisitFlorida.com)

- Questionnaires and interviews
- Domestic and International visits only; no Floridians
- Quarterly data;
- Limited origin data in state/nation resolution and no destination data

State	N visits (2018)	% of Domestic
Georgia	11935176	10.60%
New York	10021044	8.90%
Texas	4841628	4.30%
Ohio	4841628	4.30%
Pennsylvania	5742396	5.10%
Tennessee	4616436	4.10%
New Jersey	4729032	4.20%
North Carolina	5292012	4.70%
Missouri	3040092	2.70%
Illinois	5517204	4.90%
Alabama	5404608	4.80%
Maryland	2927496	2.60%
Michigan	4278648	3.80%
Virginia	3265284	2.90%
Indiana	3603072	3.20%

# Social media: data collection method



# Social media data

- 2,162,249 reviews: Floridian (42%), domestic (44%) and international (14%)
  - 250,844 reviewers visiting 51,525 properties in Florida
  - 306 origins (54 state, 67 US, 185 intern.)
  - Note: no Portuguese and Spanish language data for int. visitors.
- Metadata:
  - Reviewer's ID
  - Reviewer's home address (self-reported)
  - Reviewer's total review numbers
  - Reviewed property ID
  - Property type
  - Property location coordinates
  - Review date.

# Mobile phone data

- AirSage: trips between 11,442 zones, October 2018 and September 2019.
- A consecutive period of up to 12-months, one report per month
- Aggregation: Average weekday (Monday-Thursday), average Friday, average Saturday, average Sunday presented for the study months
- Trip Purpose Designation: Home, Work, and Other

## Sample data

Origin_Zone	Dest_Zone	Home_Zone	Month	Aggregation	Purpose	Time_of_Day	Count
10010201001	10010201001	10010201001	201802	Mon_Tue_Wed_Thu	OO	H13:H14	0.9486
10010201001	10010201001	10010201001	201802	Sat	OO	H13:H14	2.4323
10010201001	10010201001	10010201002	201802	Mon_Tue_Wed_Thu	OO	H21:H22	0.5274



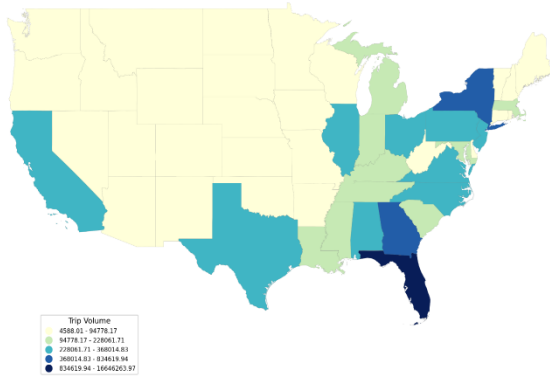
# OD matrix construction

	Segment	Origin	Destination	Geo Resolution
Social media data	Floridian	Yes	Yes	County - County
	Domestic	Yes	Yes	State - County
	Int'l	Yes	Yes	Nation - County
Cellphone data	Floridian	Yes	Yes	Tract - Tract
	Domestic	Yes	Yes	State
	Int'l	No	No	-
Survey data	Floridian	No	No	-
	Domestic	Yes	No	State
	Int'l	Yes	No	Nation

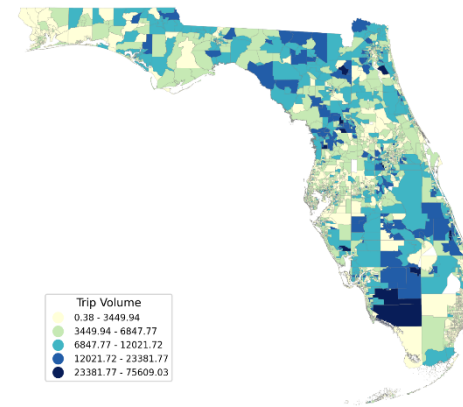
	OD /DO matrix	DD matrix
<b>Floridians</b>	county*county level (67*67 matrix) tract*tract level (3458*3458 matrix)	county*county level (67*67 matrix) tract *tract level (3458*3458 matrix)
<b>Domestic</b>	state * county level (54*67 matrix)	county*county level (67*67 matrix)
<b>International</b>	nation * county level (185*67 matrices)	county*county level (67*67 matrix)

# Data review: traveler origins

## US travelers



## Florida travelers



#	state	# of review	# of trips
1	NY	82422	34242
2	GA	71622	30639
3	PA	54314	19783
4	OH	50664	18393
5	IL	48794	18124
6	TX	42342	16726
7	NJ	43636	16661
8	NC	41422	15850
9	MA	36929	14162
10	MI	38288	13589
11	TN	35180	12853
12	VA	32912	12605

#	Origin	Name	# of Trips	Population
1	12099	Palm beach	15309	1320134
2	12011	Broward	15048	1748066
3	12095	Orange	14447	1145956
4	12057	Hillsborough	13325	1229226
5	12086	Miami-dade	12986	2496435
6	12103	Pinellas	10156	916542
7	12115	Sarasota	9578	379448
8	12071	Lee	9218	618754
9	12031	Duval	8711	864263
10	12009	Brevard	6898	543376
11	12127	Volusia	4724	494593
12	12105	Polk	4618	602095

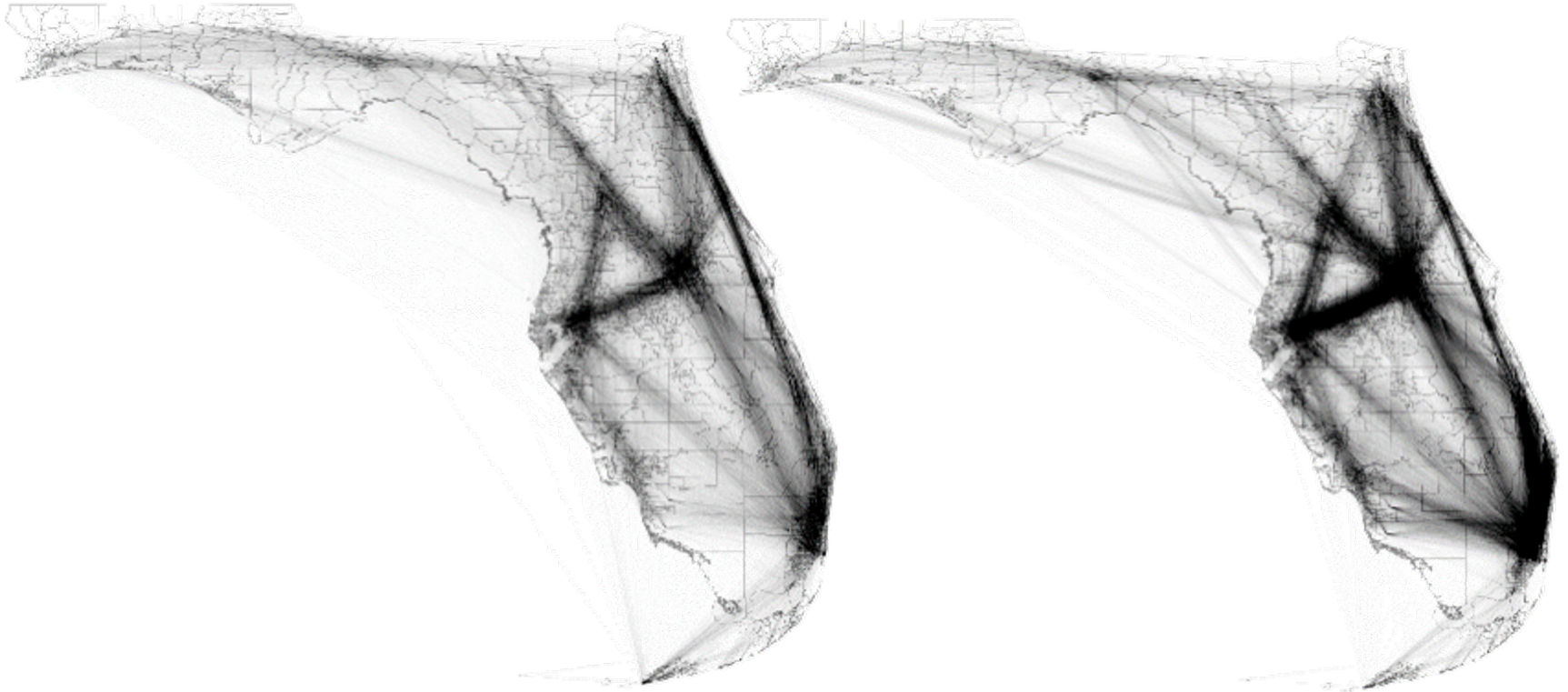
# Traveler destinations. Red color: 50% visits

County	General area	Internat.	Domestic	Floridian
Orange	Orlando	39.2%	23.0%	17.0%
Miami-Dade	Miami	20.6%	9.7%	5.2%
Broward	Fort Lauderdale	7.5%	7.1%	4.3%
Monroe	Everglades	3.3%	6.4%	7.5%
Pinellas	St. Peter	4.8%	6.7%	5.9%
Hillsborough	Tampa	2.4%	4.1%	4.7%
Lee	Fort Myers	1.9%	4.1%	4.7%
Palm beach	Palm beach	2.4%	4.2%	3.8%
Collier	Naples	2.7%	3.2%	2.9%
St. Johns	St. Augustine	0.9%	2.2%	5.3%
Volusia	Daytona	1.2%	2.8%	3.8%
Osceola	Kissimmee	3.5%	2.2%	3.0%
Sarasota	Sarasota	1.8%	2.7%	3.0%
Brevard	Cape Canaveral	1.4%	2.4%	2.6%
Duval	Jacksonville	0.6%	2.2%	3.1%
Bay	Panama City	0.3%	2.0%	0.9%
Manatee	Manatee county	1.0%	1.3%	1.2%
Okaloosa	Destin	0.2%	1.6%	0.9%
Escambia	Pensacola	0.2%	1.4%	0.9%
Polk	Lakeland	0.7%	0.8%	1.6%
Alachua	Gainesville	0.2%	0.5%	2.0%
Leon	Tallahassee	0.2%	0.5%	1.9%

# Travel network (top 5% of most frequent trips)

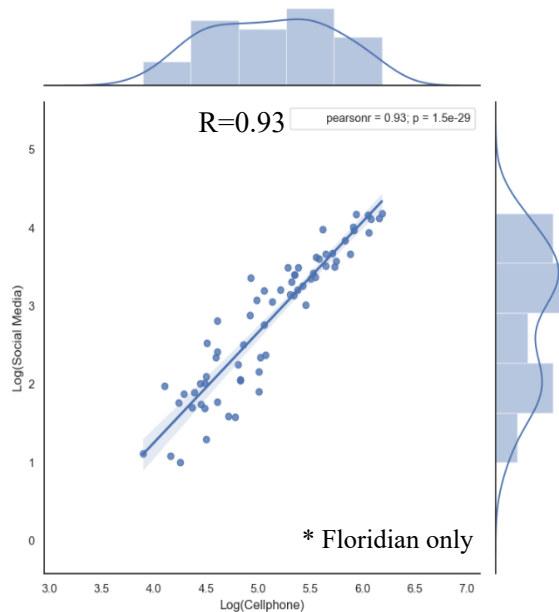
Domestic visitors

Florida travelers

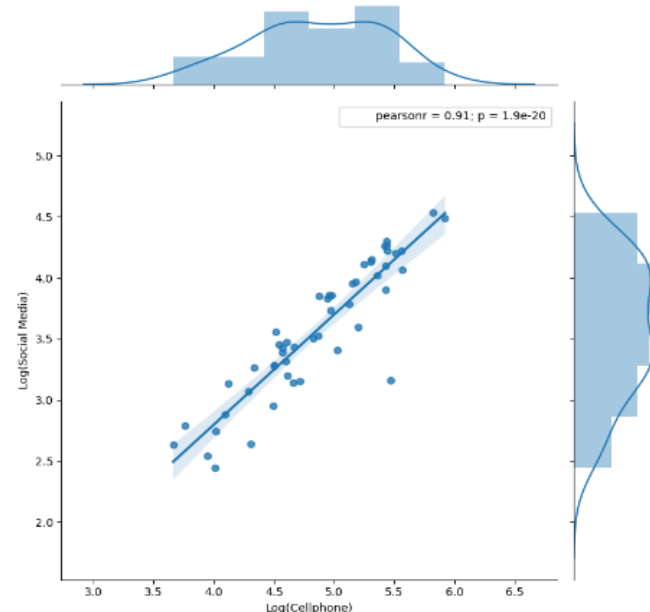


# Cross-validation of origin data

Group	Data source	Resolution	Time Aggregate
Floridian	Social Media * Cellphone	County	Overall
Domestic	Social Media * Cellphone * Survey	State	Overall



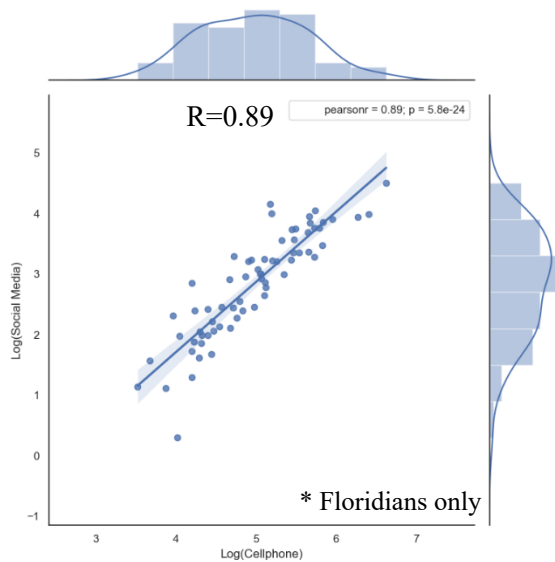
Floridians:  
 Social Media \* Cellphone  
 R=0.93



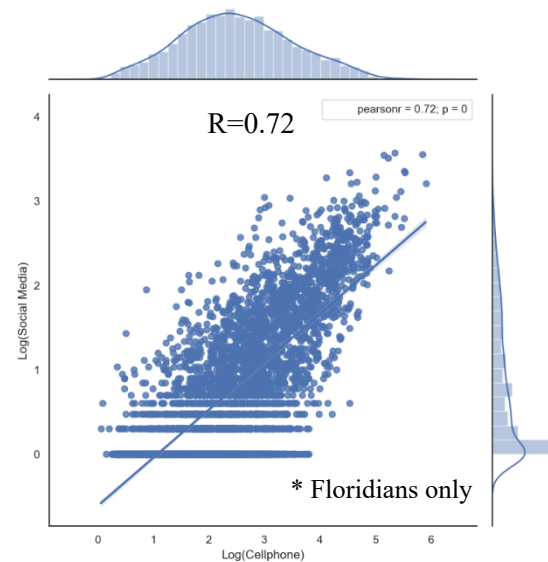
Domestic:  
 Social Media \* Cellphone R=0.91  
 Social Media \* Survey: R=0.81;  
 Cellphone \* Survey: R=0.86

# Cross-validation of destination and travel network data (Floridians)

Group	Data source	Resolution	Time Aggregate
destination	Social Media * Cellphone	County	Overall
Travel network	Social Media * Cellphone	County - County	Overall



**Destination:  $R=0.89$**



**Network:  $R=0.72$**

# Conclusions

- Methodology developed to include tourist data into provisional transportation model
- Successful validation of data coming from different sources for domestic tourists
- Significant differences between Floridian, domestic, and international travelers
- International tourists: there is a large segment of Portuguese- and Spanish- speaking tourists which show significant variation between years
  - not expected in the original proposal, data need additional validation and calibration prior to be useful



Visitors pack Anhui province's Huangshan mountain park on April 4, exceeding the visitor limit of 20,000.





# Questions?

