



AIS for Improving Maritime Statistics

Marcus Justesen, Statistics Sweden
NFGS OSLO 2017



facebook.com/statistiskacentralbyranscb



[@SCB_nyheter](https://twitter.com/SCB_nyheter)



[statistiska_centralbyran_scb](https://www.instagram.com/statistiska_centralbyran_scb)



www.linkedin.com/company/scb



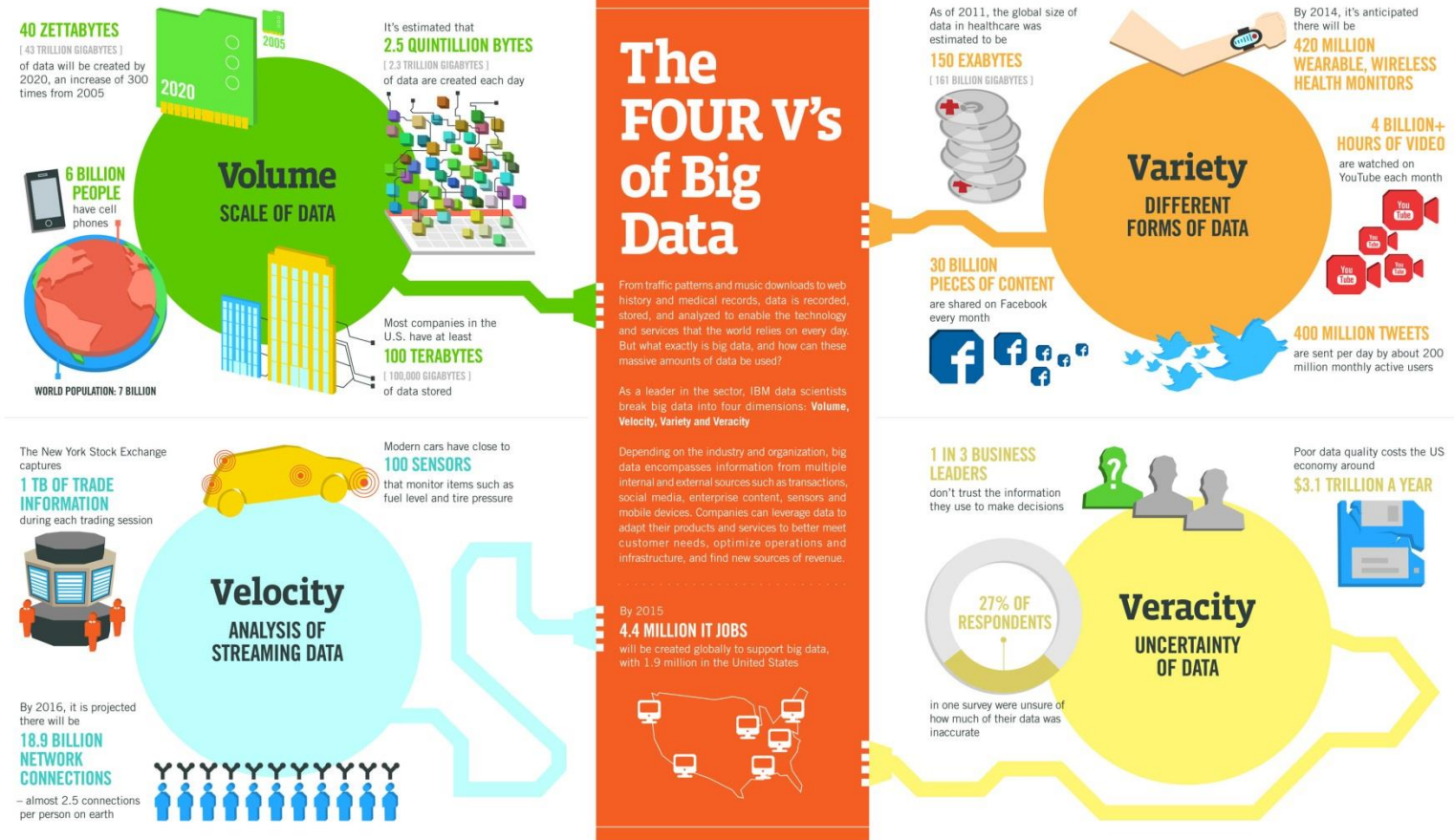


Who are doing what? And why?

THE PROJECT

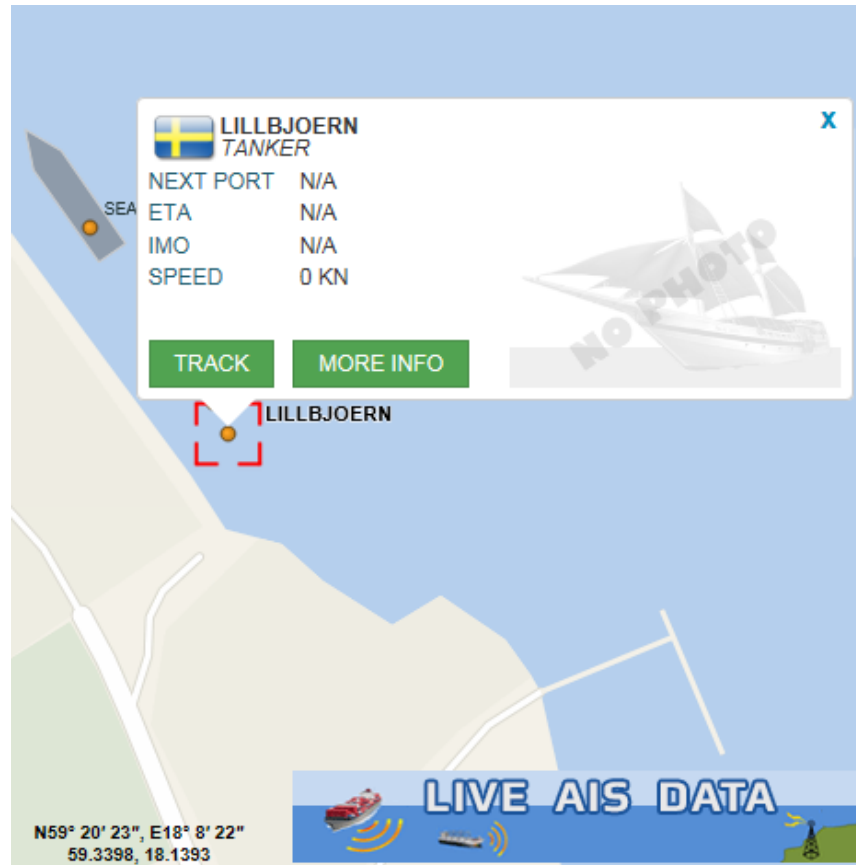


BIG DATA?



Big data source: AIS (Automatic Identification System)

- Obligatory safety system for marine trade traffic through SOLAS convention
- Digital information signals through VHF radio with position, time of signal, speed, course, type of ship etc.
- Signals can be received by any ship or land stations equipped with ais transponder
- Just a few seconds between signals
- All in all simplified: the gps system of maritime traffic



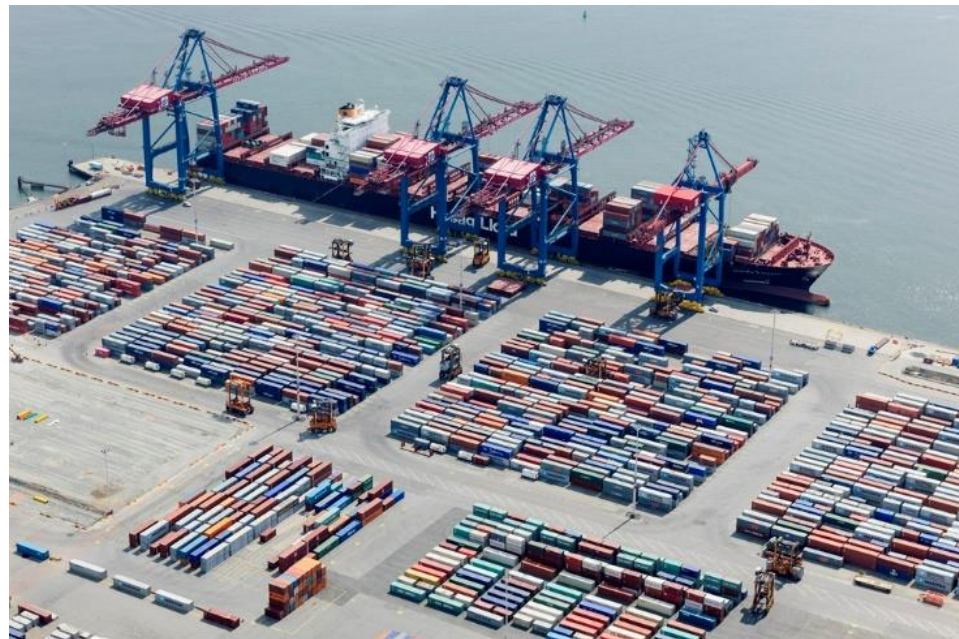
www.marinetraffic.org



Heavy burden on ports (TA:s heroes! ❤️)

Detailed information
of:

- Ship calls
- Ship
- Passangers
- vehicles
- Goods





The Objectives



- Improve quality in maritime transport statistics using AIS data.
- Evaluate opportunities for new statistics using AIS



- To gather experience working with Big Data
- Evaluate possible uses of other positional Big Data



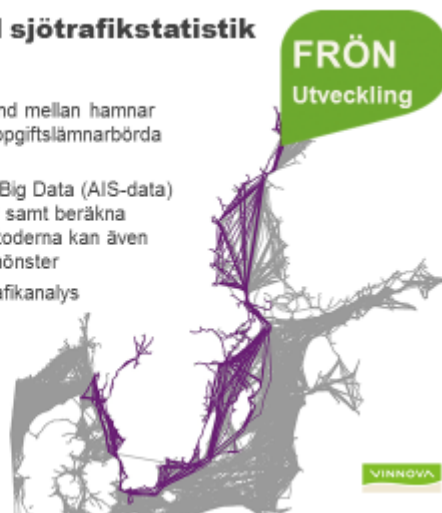
Financial support from VINNOVA (Swedens innovation agency)

Three project phases:

- Pilot
- Method development ← Reporting to Vinnova 14/6 2017
- Implementation

Metodutveckling för förbättrad sjötrafikstatistik med Big Data

- Data om fartygsanlöp till hamnar samt avstånd mellan hamnar samlas idag in via en enkät, vilket innebär uppgiftslämnarbörda samt risk för mätfel
- Utveckla metod för att använda geografiska Big Data (AIS-data) för att lista vilka fartyg som anländer hamnar samt beräkna noggrannare avstånd mellan hamnarna. Metoderna kan även användas för transportflöden och pendlingsmönster
- Statistiska centralbyrån tillsammans med Trafikanalys





Tools (100% open source)

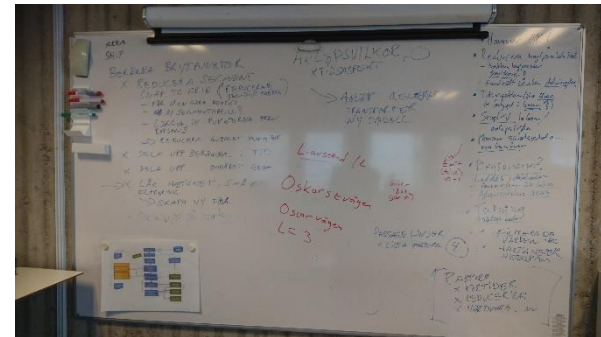
Computer:

64 GB RAM

2 TB SSD

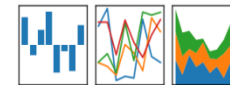
4 cores, 8 threads

Ca 700 gb data in database



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



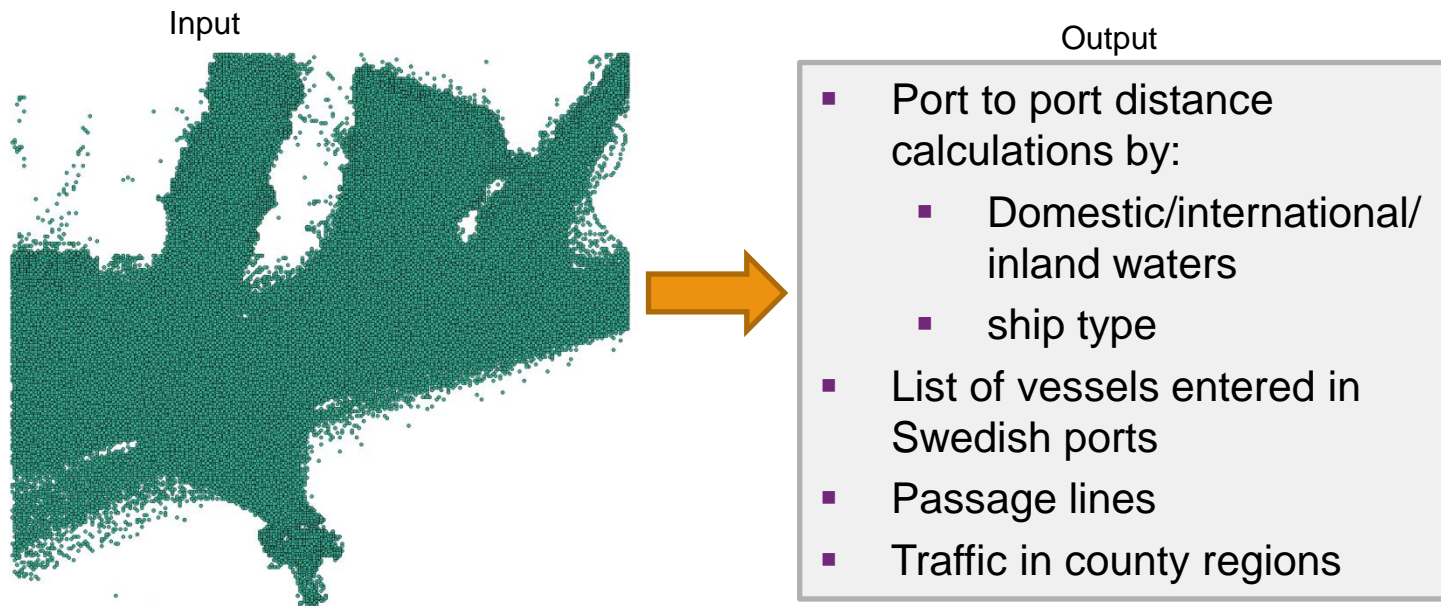


How are we doing it?

FROM BIG DATA TO DATA

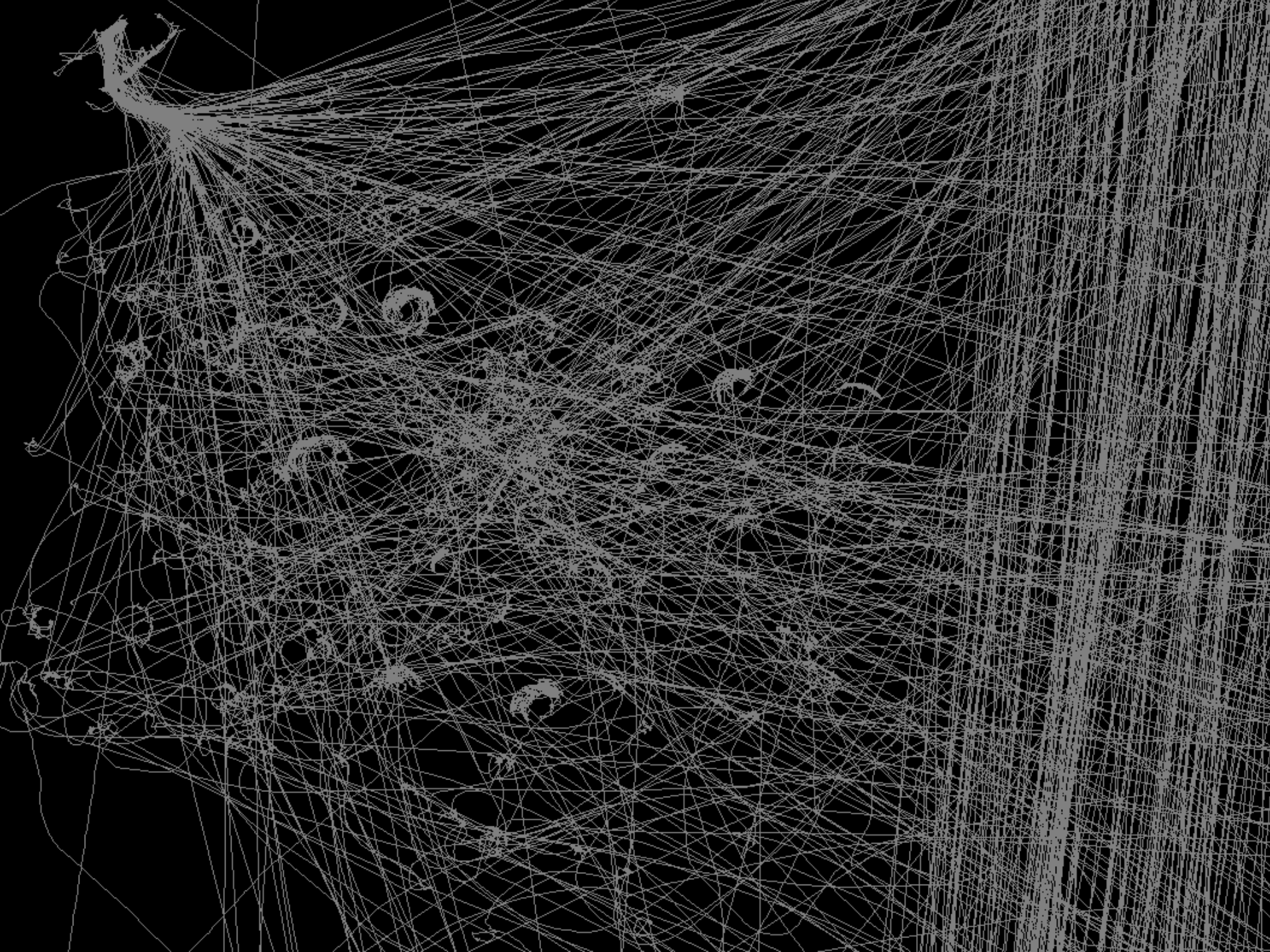


From data to statistics

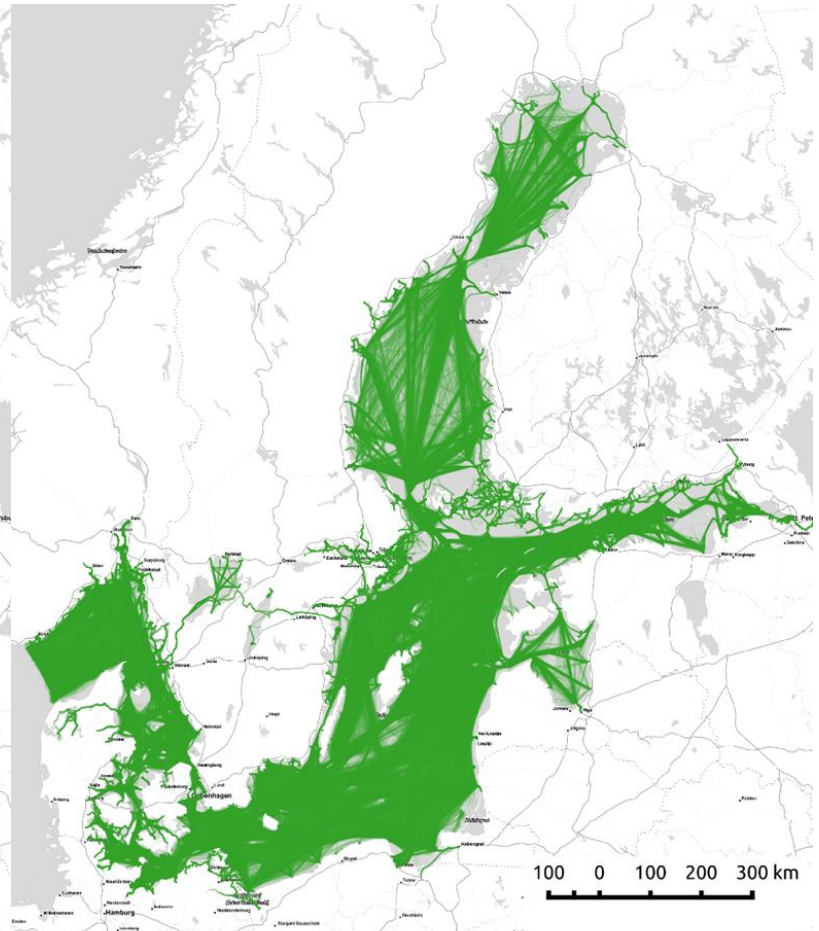
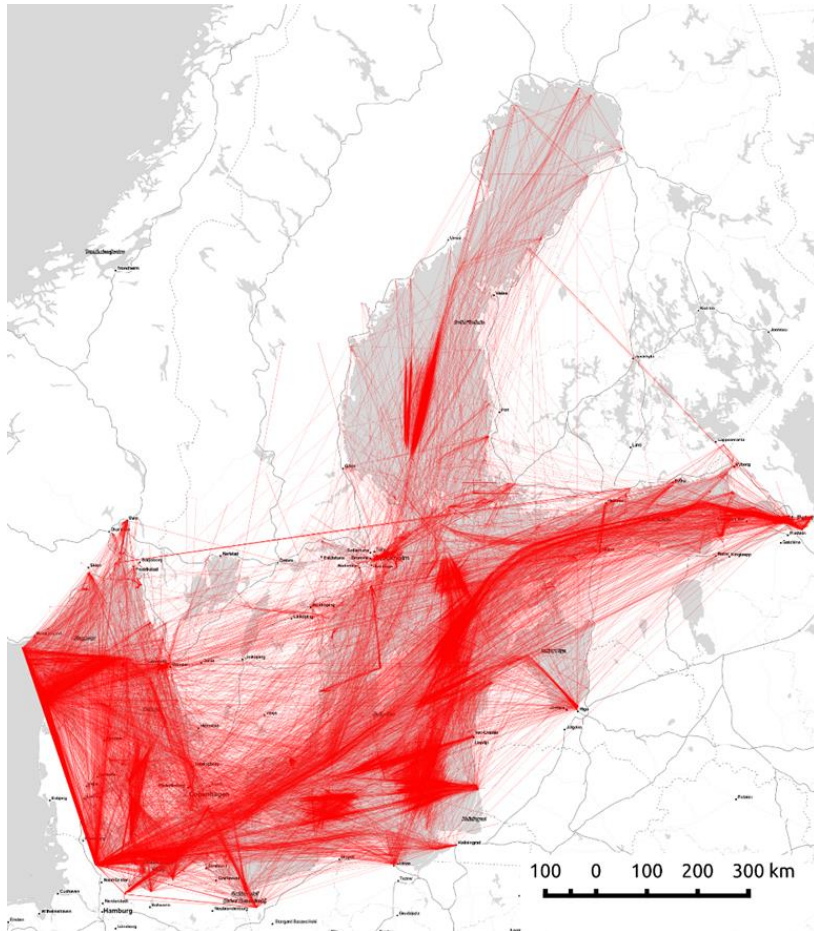


AIS data for Baltic Sea, years 2013-2015

- AIS data för Baltic Sea for 2013-2015 from Sjöfartsverket/HELCOM
 - Approximately 300 millions observations per year (330 for 2015)
 - Time interval 2-3 sec. / 6 min. (Swedish/International data) , geographical resolution of about 5-10 cm.
- Each point is a reported position with information of:
 - Ship id/transponder id, type of ship, size etc.
 - Time, speed, course etc.



Cleaning the data

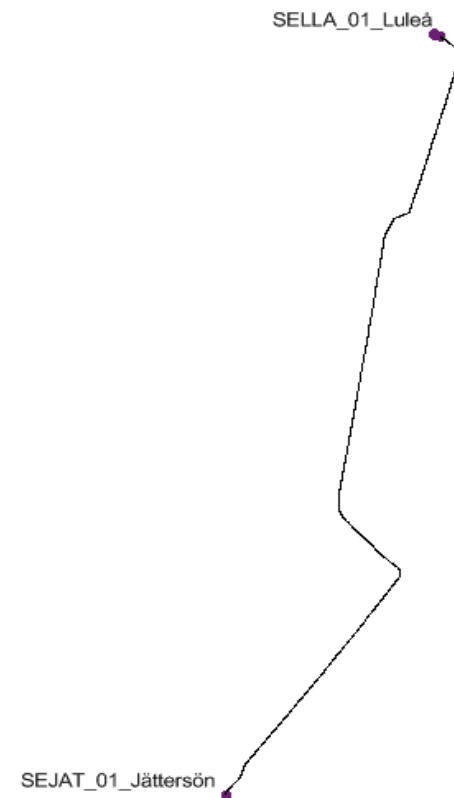




Restructuring data: creating lines

Four attributes are needed to create transport lines between ports:

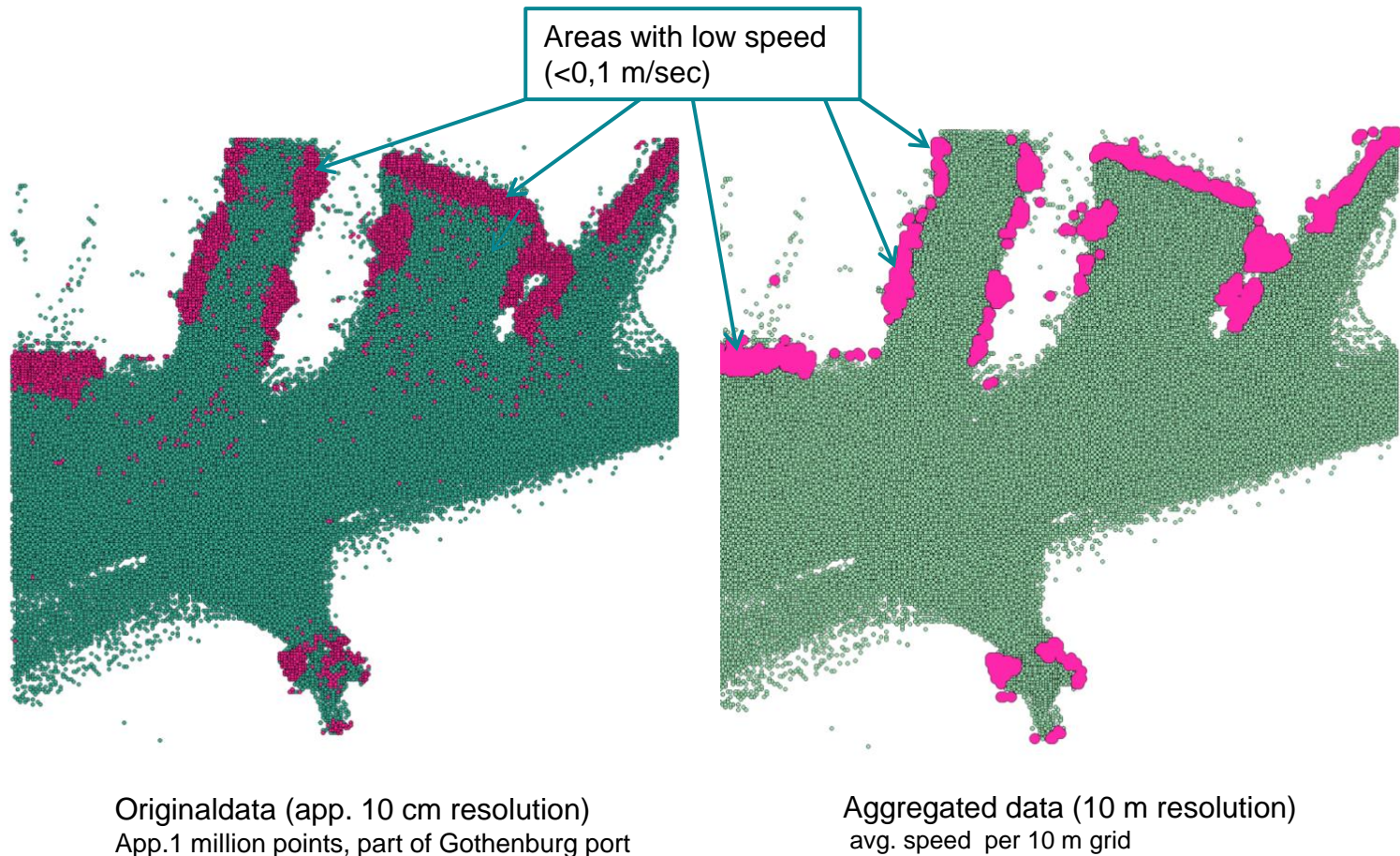
- Ship id (MMSI)
- Position
- Time
- Ports
 - This is not available to us so must be created



Berthage and ports using AIS



Average low speed = berthage

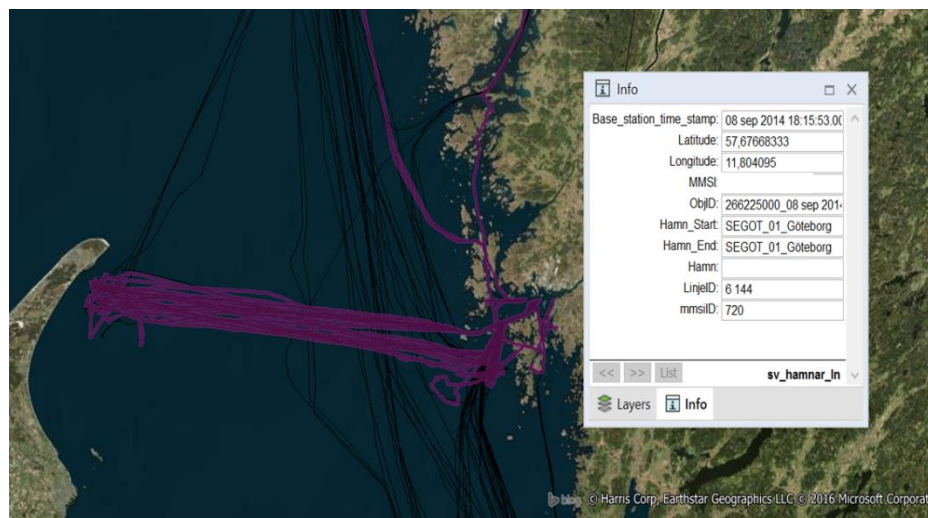


Berthing areas = hot spots of ships standing still

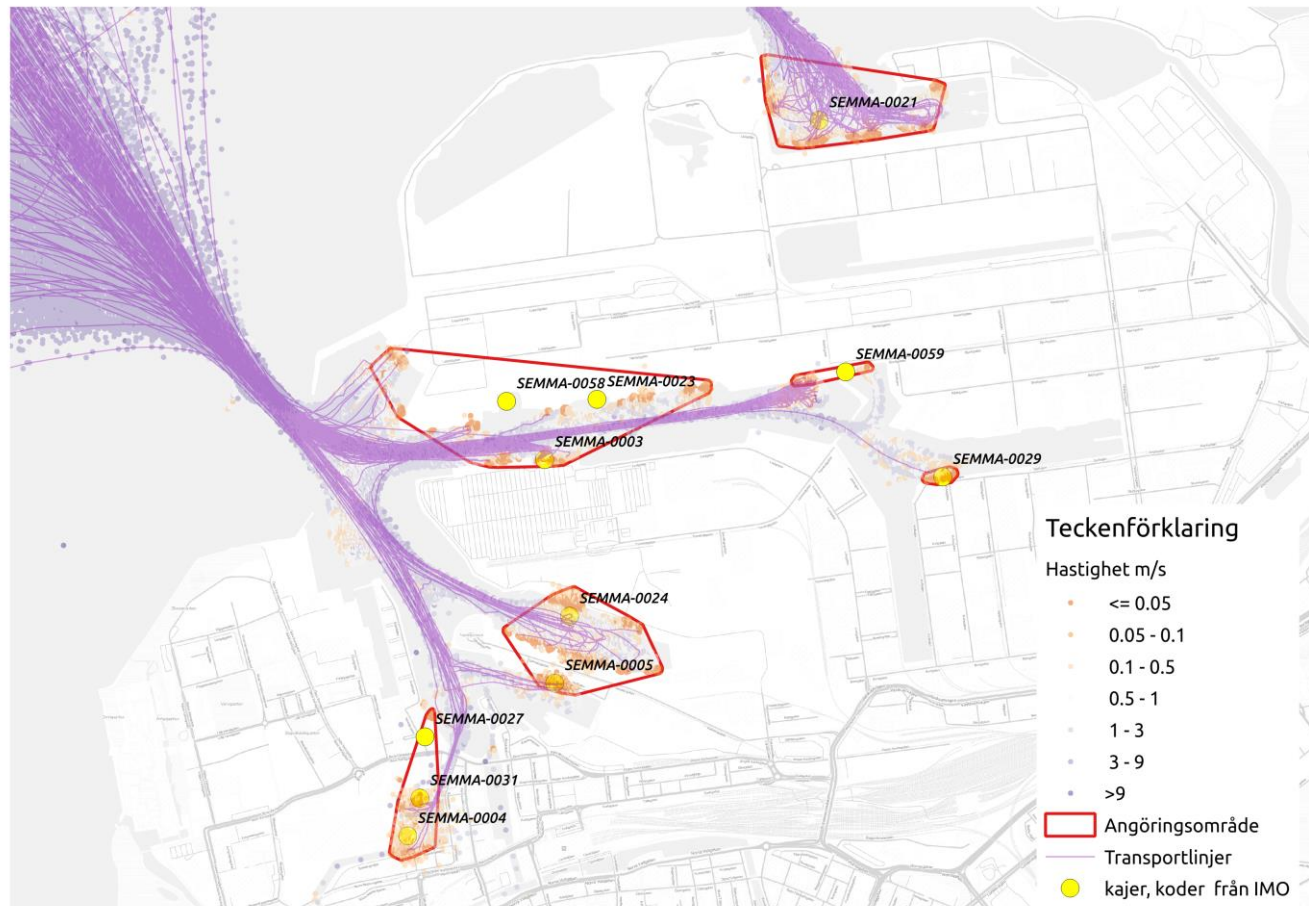


We get:

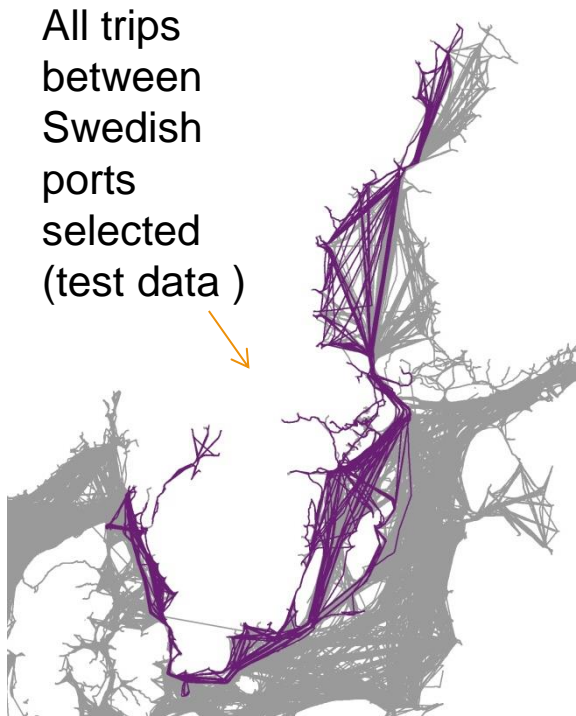
- Anchoring places
- Sluices
- Other forms of waiting places (i.e. for ice breakers)



Malmö port



Benifits of this method



- We have structured our data, making it MUCH easier to work with
 - 330 million points is now 1,3 million lines
- We now know the origin and destination of each line
- We can start producing some statistical result from the lines
- We use the lines in all further work as data source





FROM LINES TO RESULTS





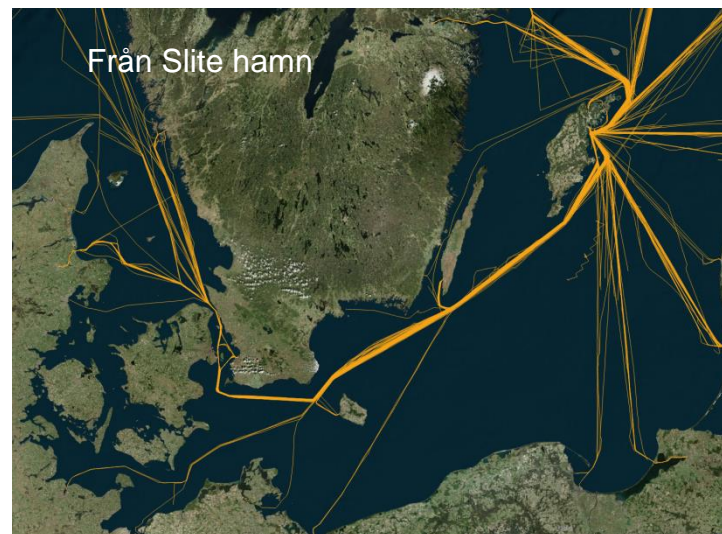
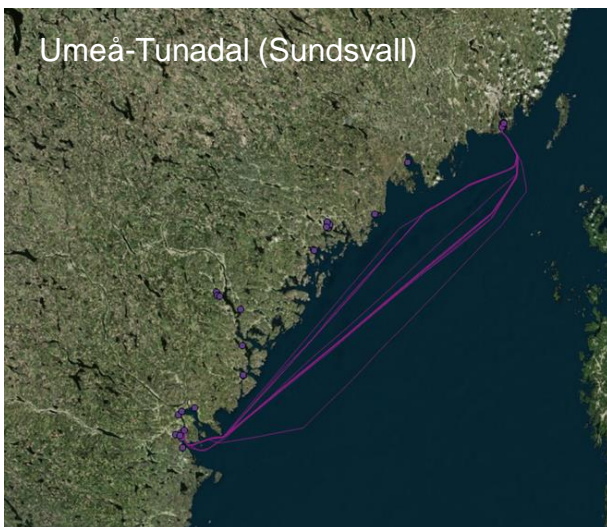
Project Goals

Develop validated methods using AIS-data for:

- Distance matrix between Swedish ports, as well as distances between Swedish and foreign ports.
- Create relevant data on ship calls to Swedish ports
- Count passages over a predefined line
- Create statistics for traffic for different type of ships and regions (such as sea basins, countys)

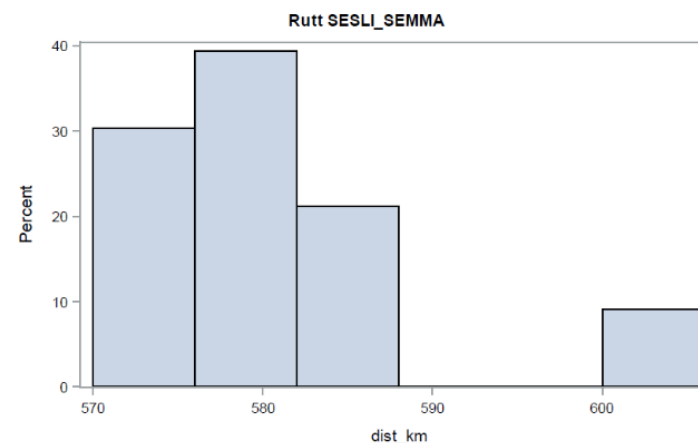
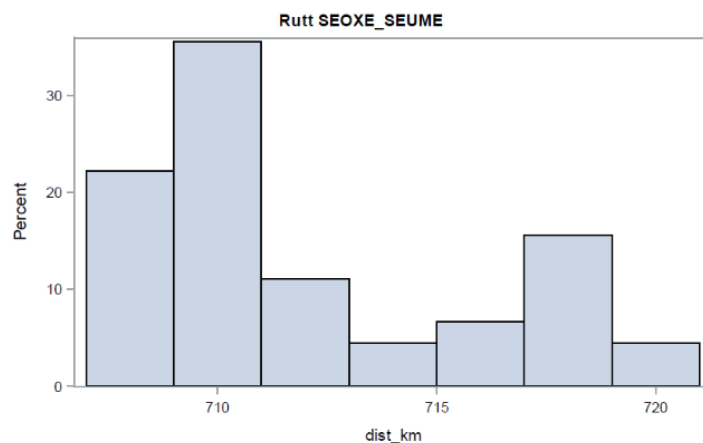
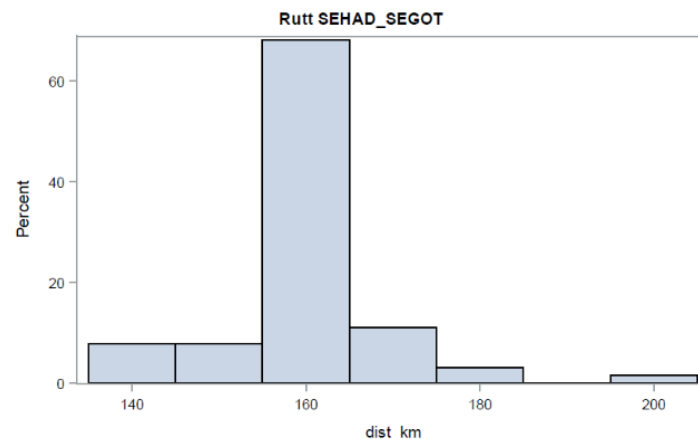
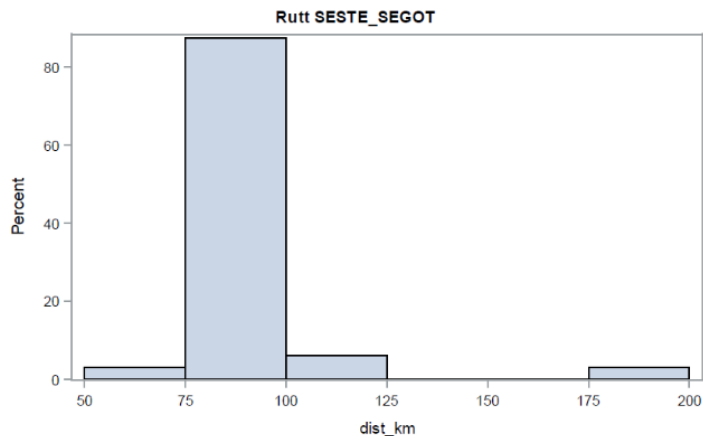


Distance matrix: one way, one distance?

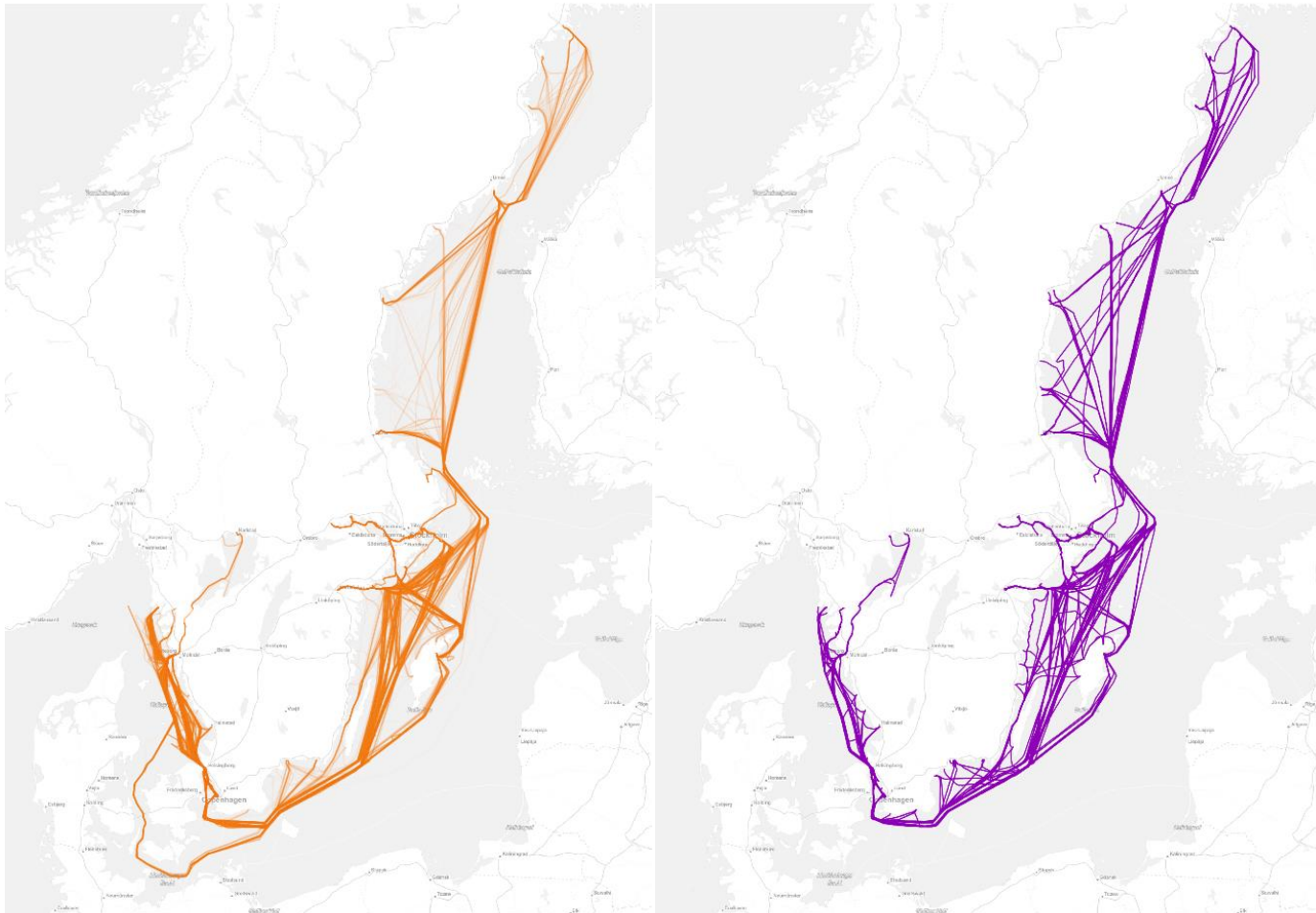




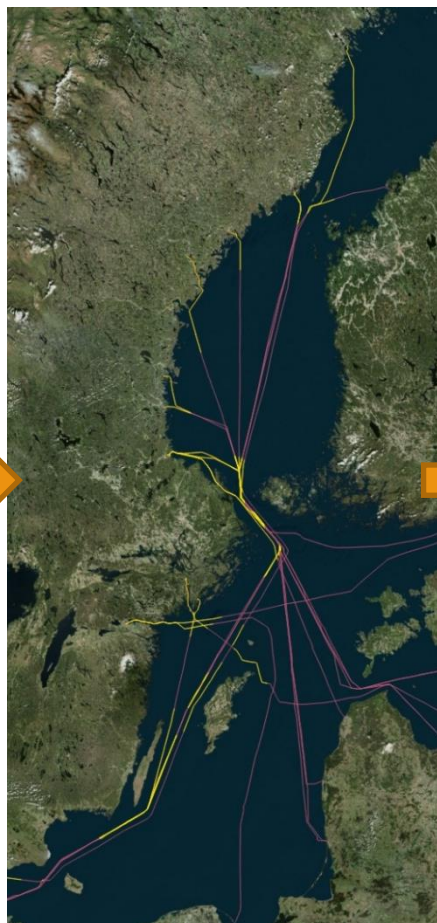
Distance variation in transports between ports



Transport lines and modeled routes in raster



Traffic in regions (Swedish waters)



Port 1	Port 2	Total Km	Km in Swedish w.
SELYS	SEGOT	114	113
SESTO	SESOE	52	51
SEVAN	SELDK	102	102
SETUN	SEIGG	125	125
SESTO	SESLI	288	241
SEGOT	SETHN	79	77
SELDK	SEVAN	101	101
SEGOT	SEHEL	213	132
SEGRU	SEVAN	144	144
SEHAK	SENRK	390	350
SEOTT	SEVAN	132	132
SEVAN	SEGRU	142	142
SEKHN	SEVAN	156	156
SEHAD	SEGOT	160	150
SEVAN	SEKHN	156	156
SEHAD	SEMMA	138	104
SEHEL	SEHAD	89	87
SENRK	SESTO	291	277
SEVAN	SEKSD	155	155
SEKSD	SEVAN	156	155
SEOXE	SE301	3	3
SE965	SETHN	0	1
SETUN	SEUME	238	197
SEVST	SE301	199	195
SESOE	SE301	102	100
SESTO	SEGVX	293	265
SEGOT	SEMMA	263	183
SETUN	SE983	1154	594



Passages



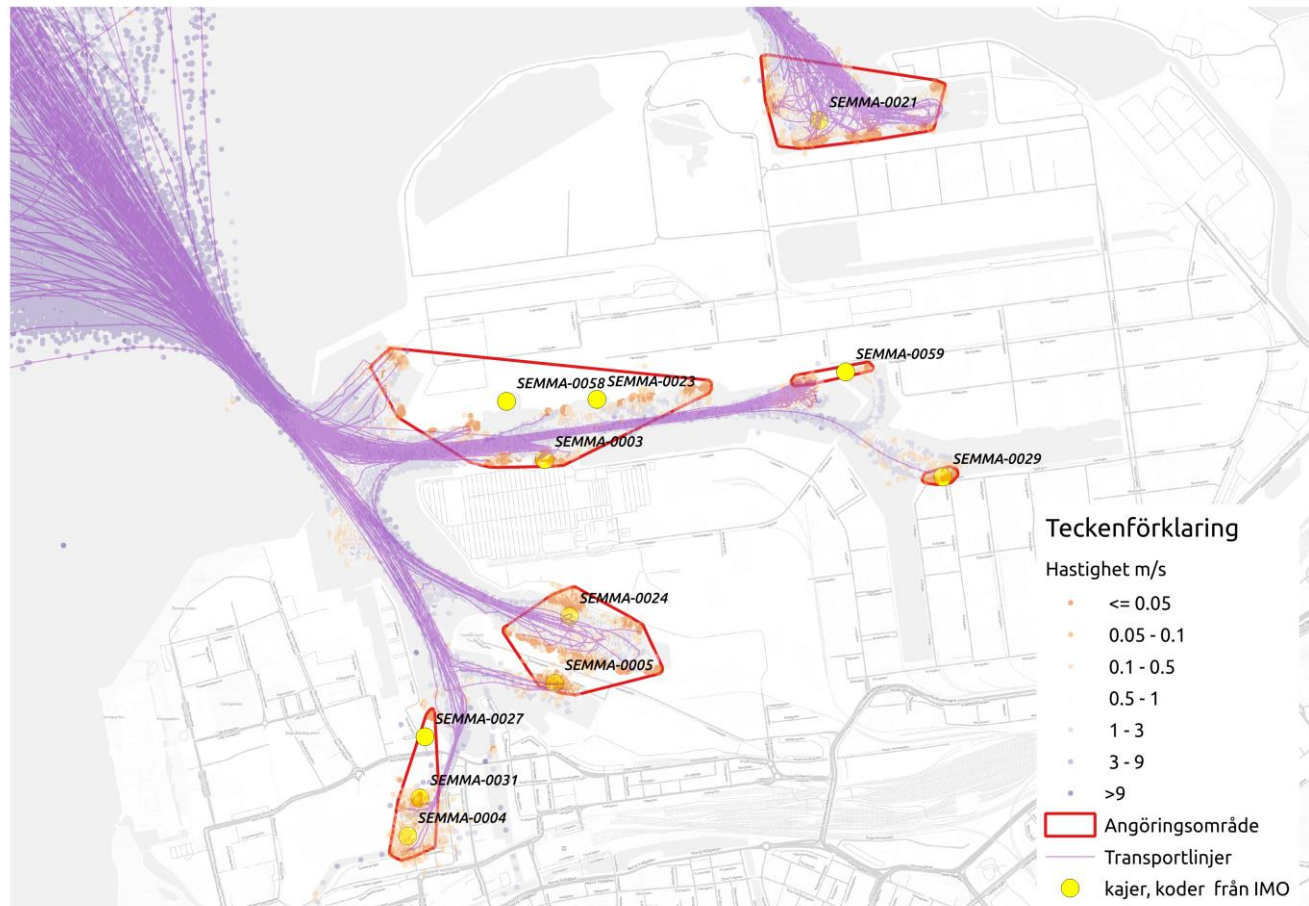
1234 = Our results

1234= Sjöfartsverket
(Swedish
Maritime Organization)

Calculations of passages
takes roughly 3 minutes!



Ship calls to ports





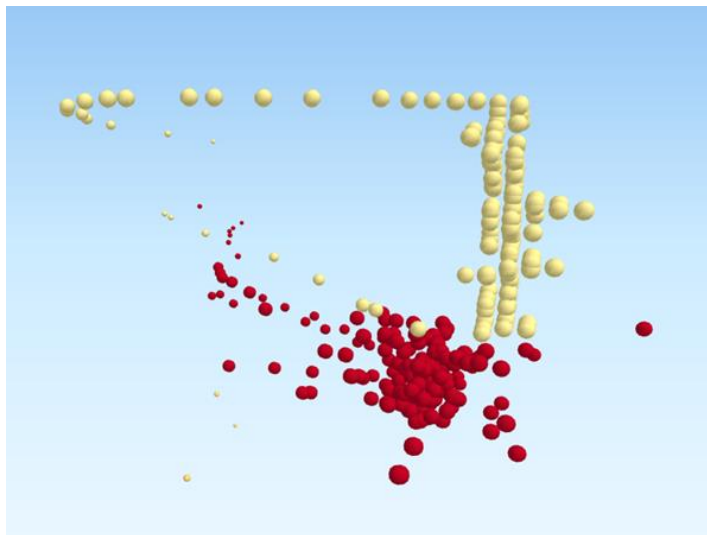
Current status

- Project finished (fine tuning of report left)
- Methods exists
- Transport Analysis has some work to do, and decisions to take based on existing results
- Up to Transport Analysis to continue to implement methods in the statistical production process





Questions?



Thank you!

Marcus Justesen | Department for Regions and Environment
Environmental Accounts and Natural Resources unit
Statistics Sweden
Box 24300, 104 51 STOCKHOLM
Visitors adress: Karlavägen 100
+46 10-479 49 61

