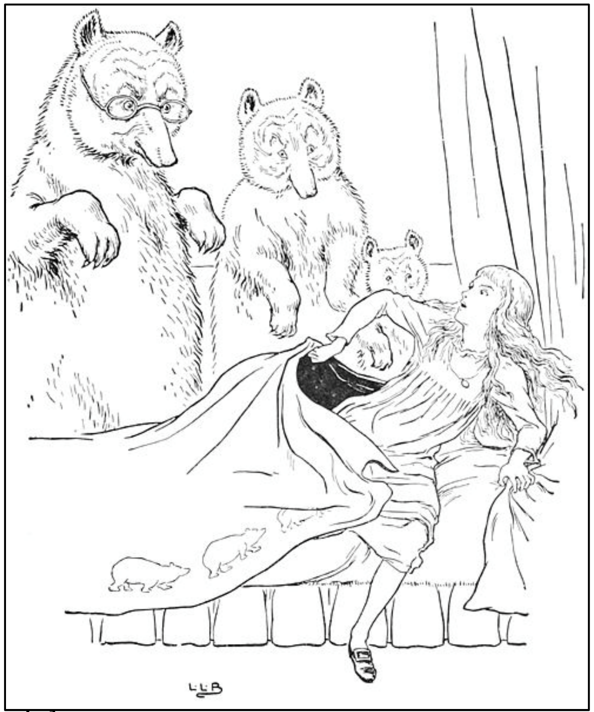
Approximation Methods

Approximation & Estimation

ap·prox·i·ma·tion ə ˌpräksəˈmāSH(ə)n/
a value or quantity that is nearly but not exactly correct.

es·ti·ma·tion estəˈmāSH(ə)n/
a rough calculation of the value, number, quantity, or extent of something. synonyms: estimate, approximation, rough calculation, rough guess, evaluation, back-of-the envelope

- Why use approximation methods?
  - Faster than more exact or precise methods,
  - Uses minimal amounts of data, and
  - Can determine if more analysis is needed:
    - Goldilocks Principle: Too big, Too little, Just right.
- Always try to estimate a solution prior to analysis.



By Leonard Leslie Brooke (archive.org) [Public domain], via Wikimedia Commons

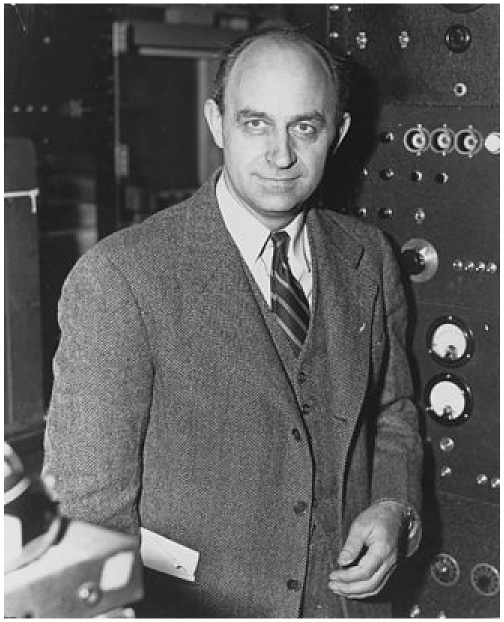
Agenda for Lesson

- Quick Examples of Approximation/Estimation
- Estimation of One-to-Many Distribution
  - Linehaul distance
  - Traveling Salesman and Vehicle Routing Problems
- Example: OfficeMin Distribution

Quick Estimation

Simple Estimation Rules

1. Break the problem into pieces that you can estimate or determine directly
2. Estimate or calculate each piece independently to within an order of magnitude
3. Combine the pieces back together paying attention to units



Enrico Fermi (1901–1954)

How many piano tuners are there in Chicago?"

- There are approximately 9,000,000 people living in Chicago.
- On average, there are two persons in each household in Chicago.
- Roughly one household in twenty has a piano that is tuned regularly.
- Pianos that are tuned regularly are tuned on average about once per year.
- It takes a piano tuner about two hours to tune a piano, including travel time.
- Each piano tuner works eight hours in a day, five days in a week, and 50 weeks in a year.

Tunings per Year = (9,000,000 ppl) ÷ (2 ppl/hh) × (1 piano/20 hh) × (1 tuning/piano/year) = 225,000
Tunings per Tuner per Year = (50 wks/yr) × (5 day/wk) × (8 hrs/day) ÷ (2 hrs to tune) = 1000
Number of Piano Tuners = (225,000 tunings per year) ÷ (1000 tunings per year per tuner) = 225

Actual Number = 290!

Problem adapted from https://en.wikipedia.org/wiki/Fermi_problem
Photo By Department of Energy. Office of Public Affairs [Public domain], via Wikimedia Commons

Example 1: Caffeine Nation!

- How many cups of coffee are drunk in the US per day?

(US Population) X (Percent Coffee Drinkers) X (Number of Cups per day)

320,000,000 people	2/3 or 66%	2 cups/day/person
3.2×10^{8}	6.7×10^{-1}	2.0×10^{0}

Recall that for scientific notation:
- Multiplication: multiply the digits and add the exponents
- Division: divide the digits and subtract the exponents

e.g., $300(2000) = 3\text{x}10^{2}\,(2\text{x}10^{3}) = 6\text{x}10^{5}$

~ 42.9×10^{7} or 4.3×10^{8}
430 million cups/day

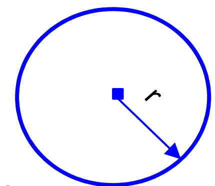
Actual Number = 400 million

source http://www.e-importz.com/coffee-statistics.php

Example 2: Next Day Distribution Centers?

- How many DCs do I need in the United States to be able to reach anywhere in the lower 48 states within a day by truck?

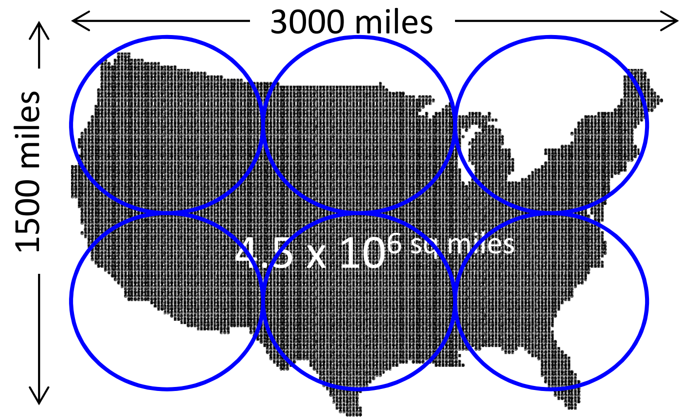
(Area of United States) $\div$ (Driving Range in One Day)



1500(3000)
$= 4.5 \times 10^6$

π (Hours to Drive/Day X Average Speed)^2

3.14 X (10 hrs/day X 50 miles/hour)^2



~ 8×10^5 sq miles

~ 0.56×10^1
~ 5.6×10^0 DCs or
6 Distribution Centers

Example 3: Selling Soda at 30,000 feet!

- How much would airlines save each year if they served a cup of soda instead of giving a full can to each passenger on domestic United States flights?

(Number of Passengers /Year) X (Savings per Passenger)

(US Population) X (flights/person/year)

(3×10^8) X (3 flights per person per year)

~9×10^8 ppl flights

(# airports) X (flights/airport-day) X (passengers/flight) X (days/year)

150 airports 100 flgts/airport-day 100 pax/flgt 365 days/yr

~6×10^8 ppl flights

(cost/can) ÷ (cups/can)

(0.2 \$/can) ÷ (2 cups/can)

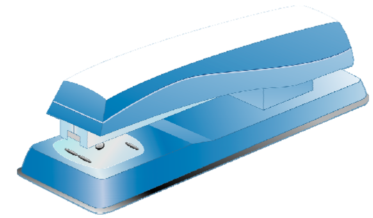
0.10 \$/cup @ 1 per passenger

~ $6\text{-}9 \times 10^7$ \$ or
~ 75 million \$

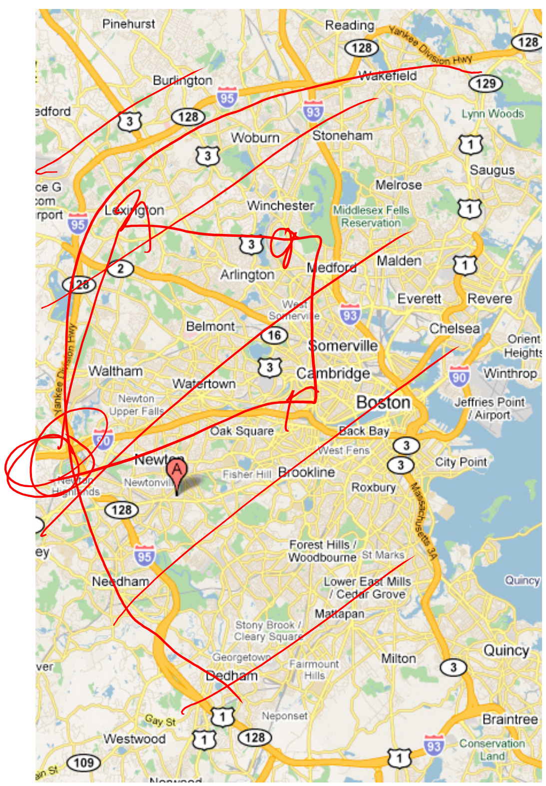
source Weinstein, L. and J. Adam, *Guestimation: Solving the World's Problems on the back of a Cocktail Napkin*.

Estimation of One to Many Distribution

Example: OfficeMin



- You have joined a new start up, OfficeMin, that wants to provide office supplies to homes and business within the greater Boston area. Initially, OfficeMin plans to deliver its office supplies to different firms within the I-95 highway loop around Boston from a distribution center located in Newton.
- The VP of distribution has asked you to find the expected cost per day for this distribution strategy.



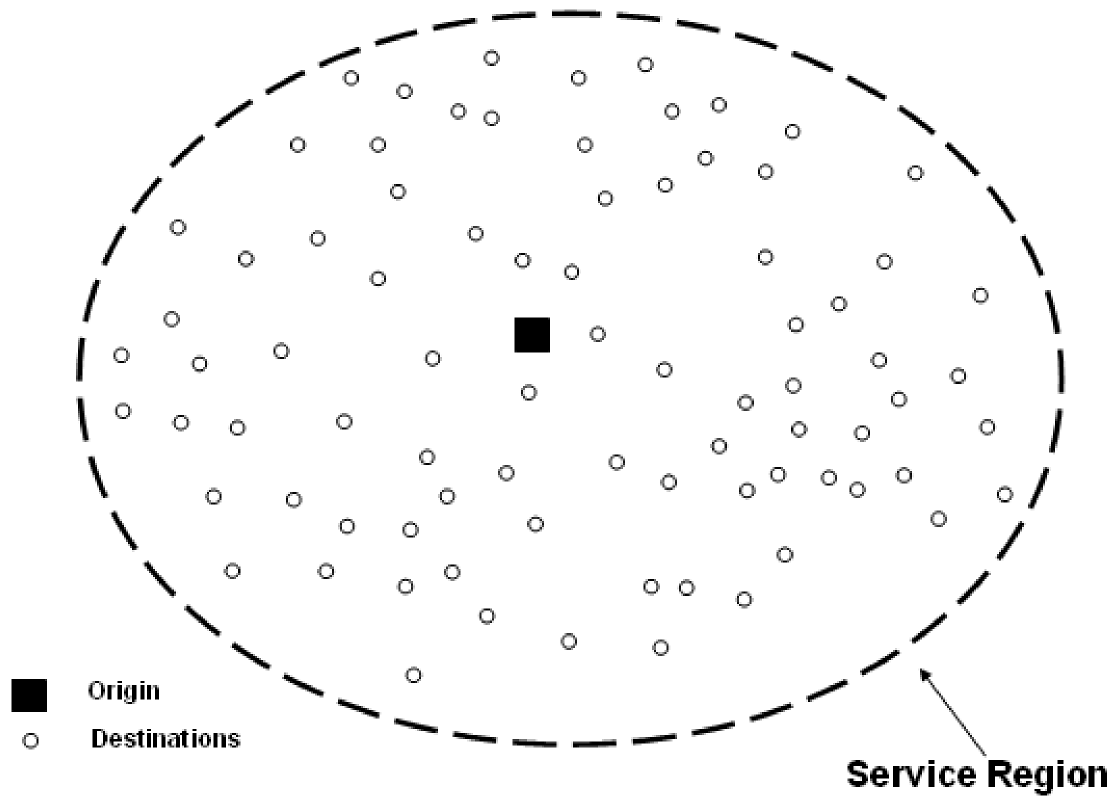
- What information do you need?
- What methodology should you use?

Image public domain from www.pixabay.com

One to Many System

Single Distribution Center:
- Products originate from one origin
- Products are demanded at many destinations
- All destinations are within a specified Service Region
- Ignore inventory (same day delivery)

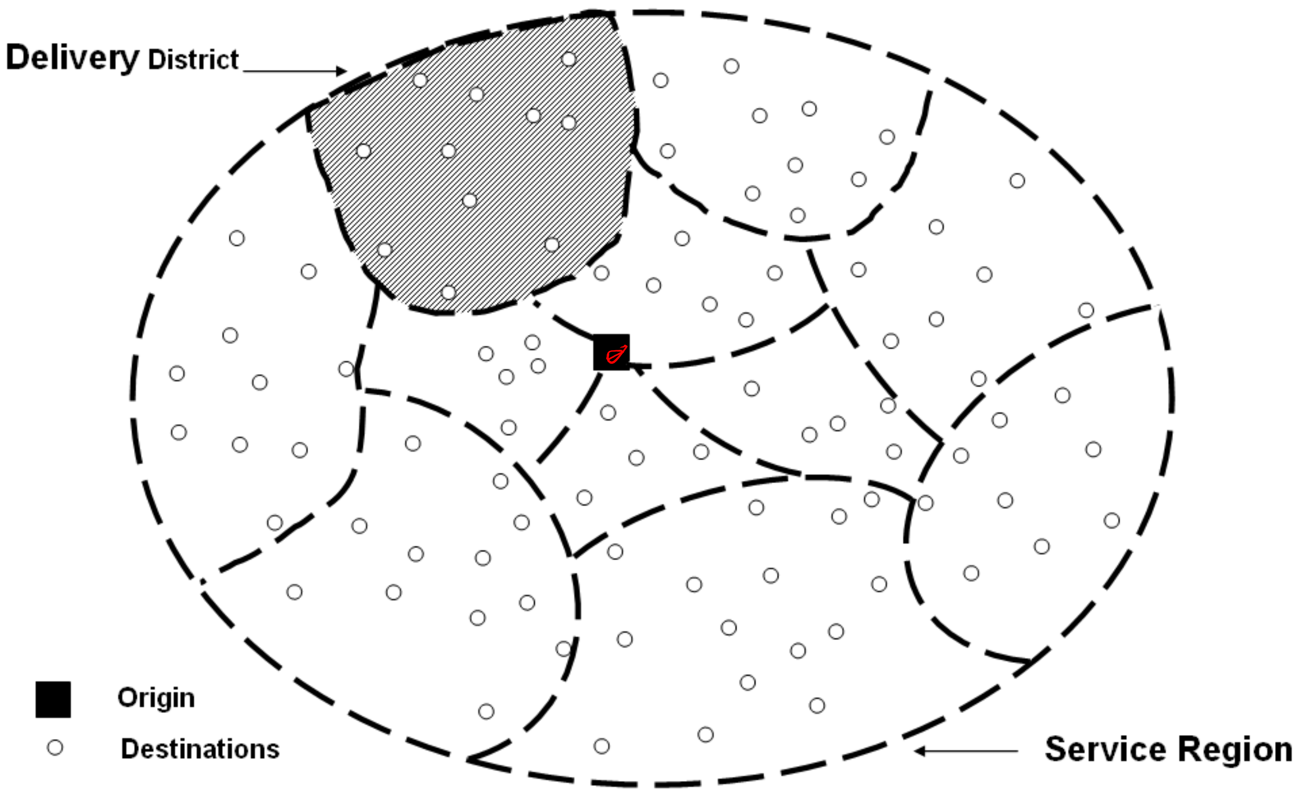
Assumptions:
- Vehicles are homogenous
- Same capacity, Q_{MAX}
- Fleet size is constant



Case adapted from Hernandez Lopez, J.J. (2003) "Evaluation of Bulk and Packaged Distribution Strategies in a Specialty Chemical Company," MIT Supply Chain Management Program Thesis.

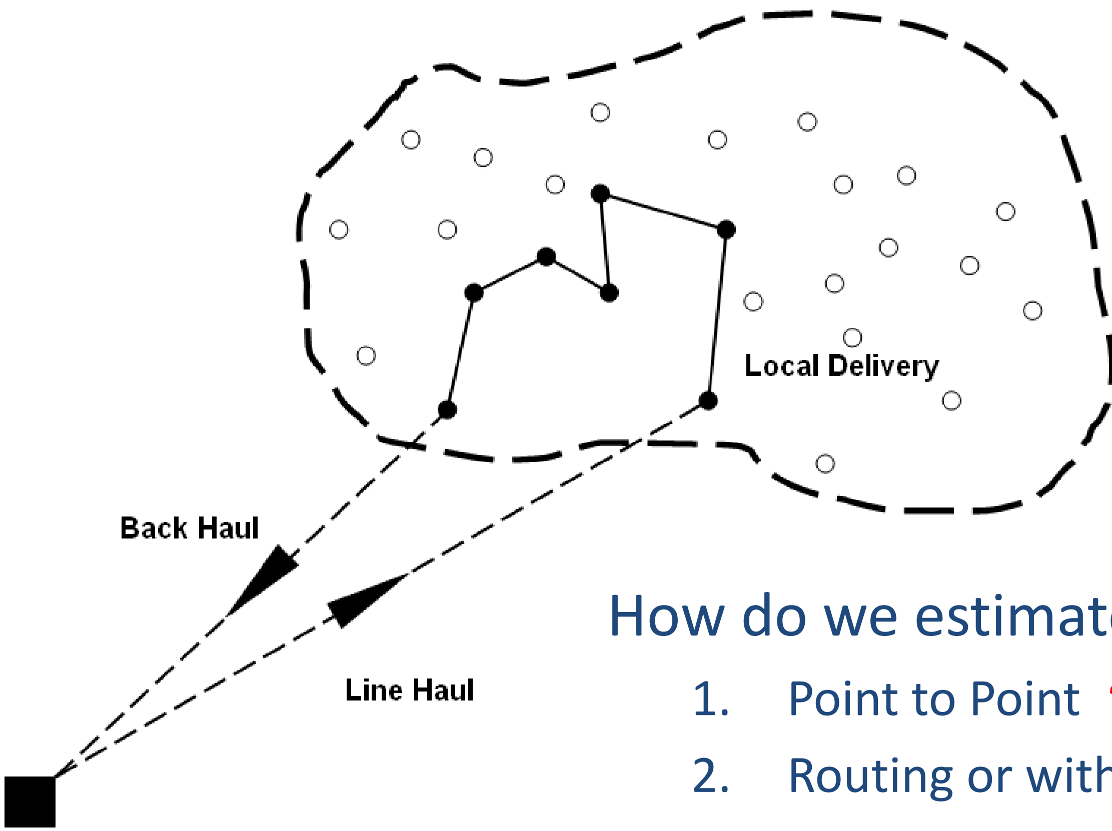
One to Many System

Finding the estimated total distance:
- Divide the Service Region into Delivery Districts
- Estimate the distance required to service each district



One to Many System

Route to serve a specific district:
- Line haul from origin to the 1st customer in the district
- Local delivery from 1st to last customer in the district
- Back haul (empty) from the last customer to the origin



$$d_{TOUR} \approx 2d_{LineHaul} + d_{Local}$$

$d_{LineHaul}$ = Distance from origin to center of gravity (centroid) of delivery district

d_{Local} = Local delivery between customers in one district

How do we estimate distances?

1. Point to Point
2. Routing or within a Tour

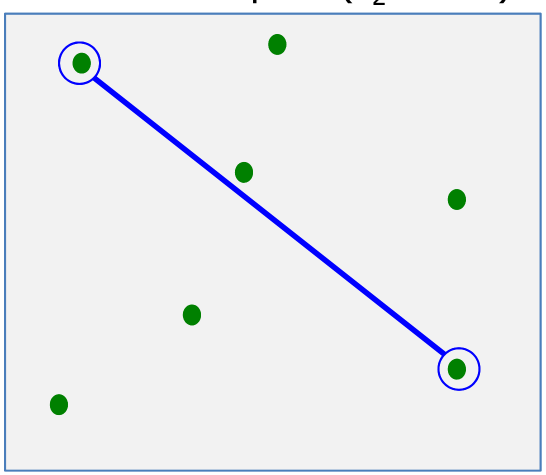
MIT Center for Transportation & Logistics

Estimating Point to Point Distances

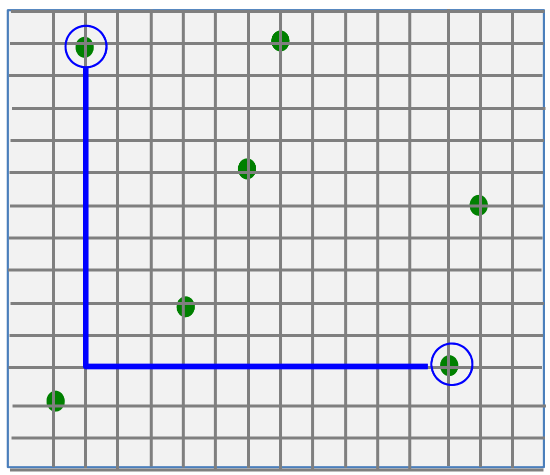
Distance Estimation: Point to Point

- Why bother?
- How to do it?
  - Depends on the topography of the underlying region
    - Euclidean Space: $d_{A\text{-}B} = \sqrt{[(x_A - x_B)^2 + (y_A - y_B)^2]}$
    - Grid: $d_{A\text{-}B} = |x_A - x_B| + |y_A - y_B|$
    - Random Network: different approach

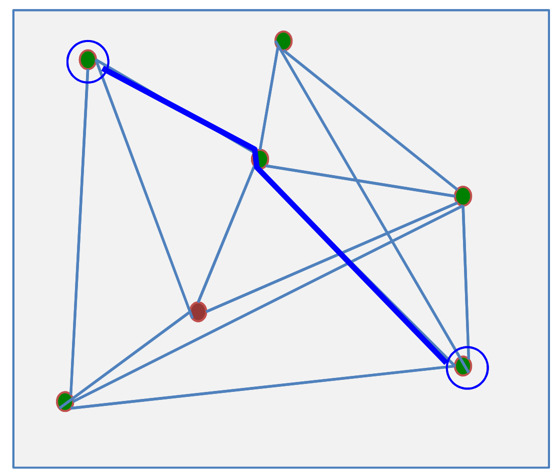
Euclidean Space (L_2 Metric)



Manhattan Metric / Grid (L_1 Metric)



Random Network



Distance Estimation: Point to Point

- For Random (real) Networks use: $D_{A\text{-}B} = k_{CF}\, d_{A\text{-}B}$
- Find $d_{A\text{-}B}$ - the "as crow flies" distance.
  - Euclidean: for really short distances
    - $d_{A\text{-}B} = SQRT((x_A - x_B)^2 + (y_A - y_B)^2)$
  - Great Circle: for locations within the same hemisphere
    - $d_{A\text{-}B} = 3959(\arccos[\sin[LAT_A]\sin[LAT_B] + \cos[LAT_A]\cos[LAT_B]\cos[LONG_A - LONG_B]])$

    Where:
    - LAT_i = Latitude of point i in radians
    - $LONG_i$ = Longitude of point i in radians
    - Radians = (Angle in Degrees)($\pi/180^o$)
- Apply an appropriate circuity factor (k_{CF})
  - How do you get this value?
  - What do you think the ranges are?
  - What are some cautions for this approach?

Selected Values of k_{CF}

Country	k_{CF}	StdDev
Argentina	1.22	0.15
Australia	1.28	0.17
Belarus	1.12	0.05
Brazil	1.23	0.11
Canada	1.30	0.10
China	1.33	0.34
Egypt	2.10	1.96
Europe	1.46	0.58
England	1.40	0.66
France	1.65	0.46
Germany	1.32	0.95
Italy	1.18	0.10
Spain	1.58	0.80
Hungary	1.35	0.25
India	1.31	0.21
Indonesia	1.43	0.34

Country	k_{CF}	StdDev
Japan	1.41	0.15
Mexico	1.46	0.43
New Zealand	2.05	1.63
Poland	1.21	0.09
Russia	1.37	0.26
Saudi Arabia	1.34	0.19
South Africa	1.23	0.12
Thailand	1.42	0.44
Turkey	1.36	0.34
Ukraine	1.29	0.12
United States	1.20	0.17
Alaska	1.79	0.87
US East	1.20	0.16
US West	1.21	0.17

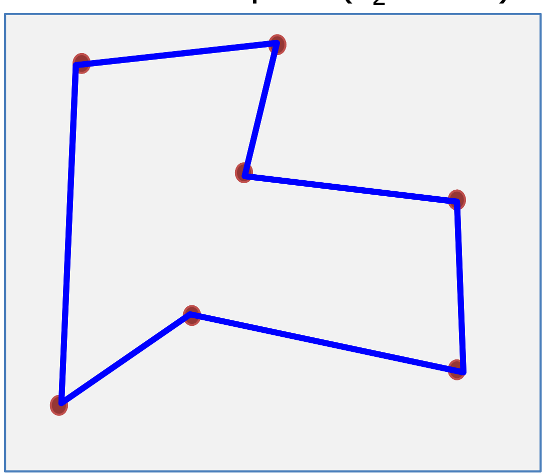
Source: Ballou, R. (2002) "*Selected country circuity factors for road travel distance estimation," Transportation Research Part A, p 843-848.*

Estimating Local Route Distances

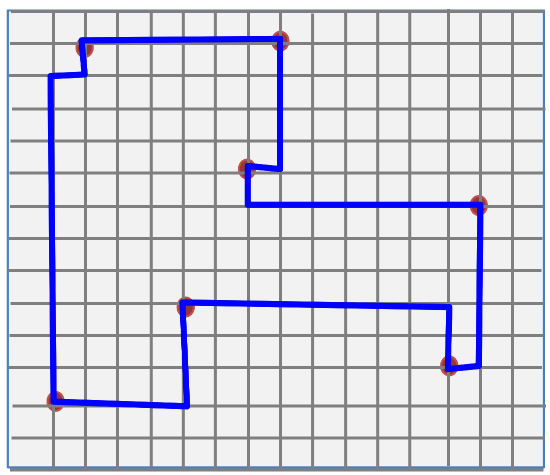
Distance Estimation: Local Routing

- Traveling Salesman Problem
  - Starting from an origin, find the minimum distance required to visit each destination once and only once and return to origin.
  - The expected TSP distance, d_{TSP}, is proportional to $\sqrt{(nA)}$ where n= number of stops and A=area of district
  - The estimation factor (k_{TSP}) is a function of the topology

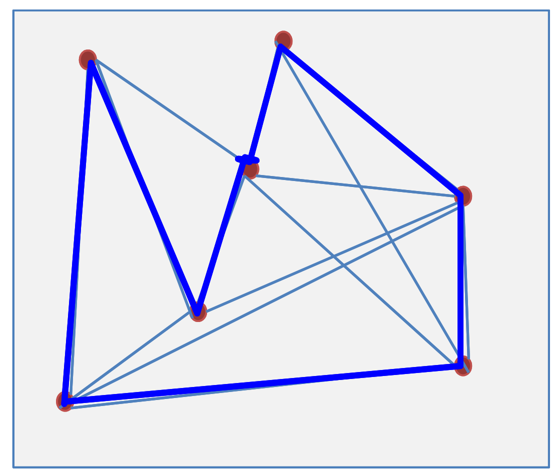
Euclidean Space (L_2 Metric)



Manhattan Metric / Grid (L_1 Metric)



Random Network



One to Many System

A = Area of district
n = Number of stops in district
δ = Density (# stops/Area)
k_{TSP} = TSP network factor (unitless)
d_{TSP}= Traveling Salesman Distance
d_{stop}= Average distance per stop

- What can we say about the expected TSP distance to cover **<u>n</u>** <u>stops</u> in district with an <u>area of **A**</u>?
- A good approximation, assuming a "fairly compact and fairly convex" area, is:

$$d_{TSP} \approx k_{TSP}\sqrt{nA} = k_{TSP}\sqrt{n\left(\frac{n}{\delta}\right)} = k_{TSP}\left(\frac{n}{\sqrt{\delta}}\right)$$

$$d_{stop} \approx \frac{k_{TSP}\sqrt{nA}}{n} = k_{TSP}\sqrt{\frac{A}{n}}$$

- What values of k_{TSP} should we use?
  - Lots of research on this for L_1 and L_2 networks - depends on district shape, approach to routing, etc.
  - Euclidean (L_2) Networks
    - k_{TSP} = 0.57 to 0.99 depending on clustering & size of N (MAPE~4%, MPE~-1%)
    - k_{TSP}=0.765 commonly used and is a good approximation!
  - Grid (L_1) Networks
    - k_{TSP} = 0.97 to 1.15 depending on clustering and partitioning of district

References: Daganzo, C.. (2010) <u>Logistics Systems Analysis</u>, 4th Edition Springer-Verlag.
Larson, R. and Odoni, A. (1981) <u>Urban Operations Research</u>, http://web.mit.edu/urban_or_book/www/book/

Estimating Vehicle Tour Distances

Estimating Total Tour (VRP) Distance

- Finding the total distance traveled on all tours, where:
  - l = number of tours
  - c = number of customer stops per tour and
  - n=total number of stops = c*l

$$d_{TOUR} = 2d_{LineHaul} + \frac{ck_{TSP}}{\sqrt{\delta}}$$

$$d_{AllTours} = ld_{TOUR} = 2ld_{LineHaul} + \frac{nk_{TSP}}{\sqrt{\delta}}$$

- Minimize number of tours by maximizing vehicle capacity

$$l = \left[\frac{D}{Q_{MAX}}\right]^{+}$$

$$d_{AllTours} = 2\left[\frac{D}{Q_{MAX}}\right]^{+} d_{LineHaul} + \frac{nk_{TSP}}{\sqrt{\delta}}$$

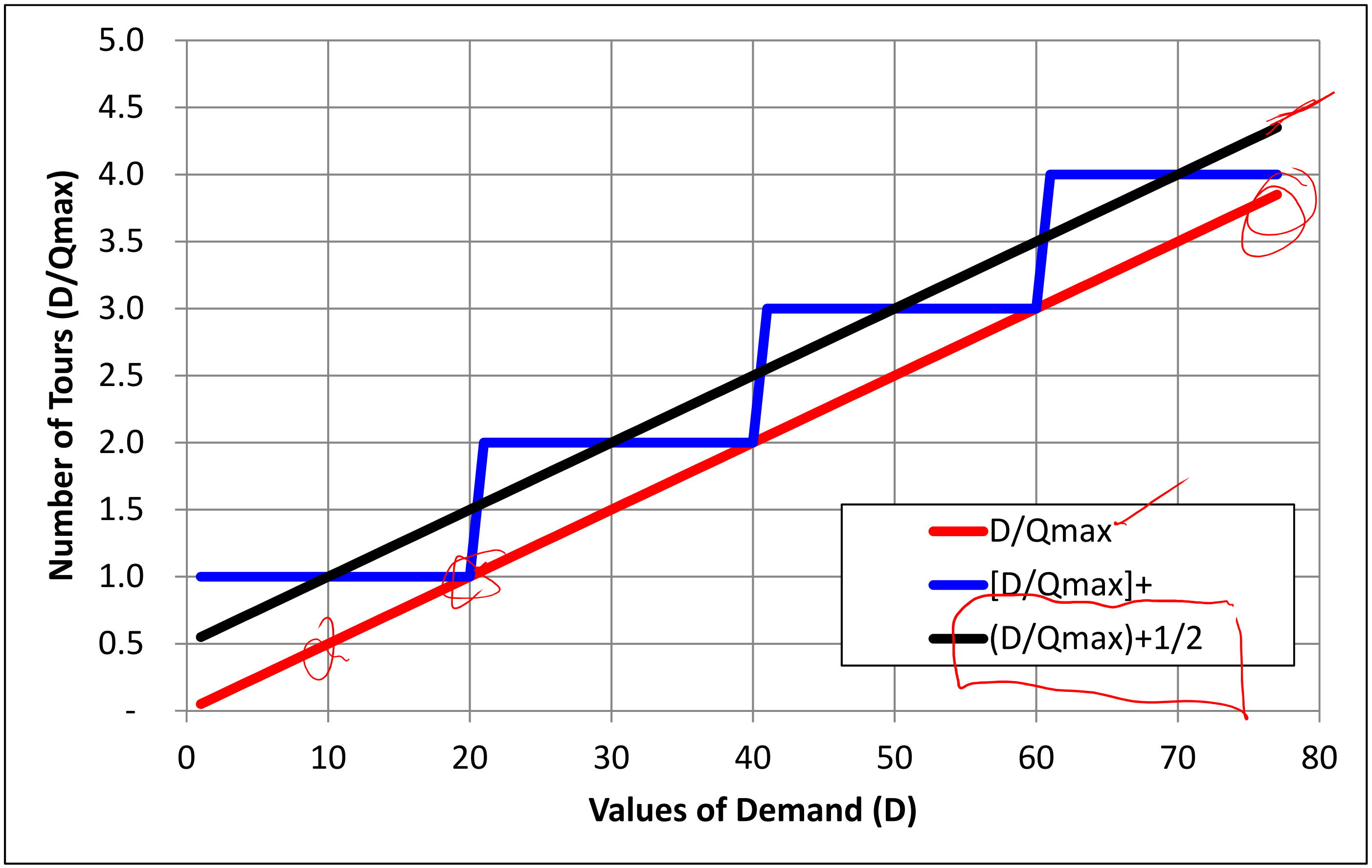
$[x]^{+}$ = lowest integer value > x. This is a step function

Estimate this with continuous function:

$[x]^{+} \sim x + ½$

Continuous Approximation

In this example, Q_{MAX}=20. The number of tours, l, would be $[D/Q_{MAX}]^+$ which is a step function. Step functions are not continuous – lets create a continuous approximation of this function that we can use.



Putting it all together

The approximate transportation cost to deliver to all customers becomes:

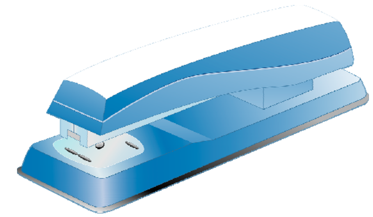
$$TransportCost = c_s \overbrace{\left[n + \frac{D}{Q_{MAX}} + \frac{1}{2} \right]}^{\text{Expected number of loads/unloads (customer + origin)}} + c_d \left(2 \overbrace{\left[\frac{D}{Q_{MAX}} + \frac{1}{2} \right]}^{\text{Expected number of linehaul moves or tours}} \overbrace{d_{LineHaul}}^{\text{Expected distance for the linehaul portion. How do I get this?}} + \overbrace{\frac{nk_{TSP}}{\sqrt{\delta}}}^{\text{Expected local delivery distance regardless of number of tours}} \right) + \underbrace{c_{vs}}_{\text{Cost per item per stop}} D$$

n =Expected number of stops in district
D =Expected demand in district
Q_{MAX}=Capacity of each truck
c_s=Cost per stop ($/stop)
c_d=Cost per distance ($/mile)
c_{vs}=Cost per unit per stop ($/item-stop)

δ=Density (# stops/Area)
k_{TSP} =TSP network factor (unitless)
d_{TSP}=Traveling Salesman Distance
d_{stop}=Average distance per stop

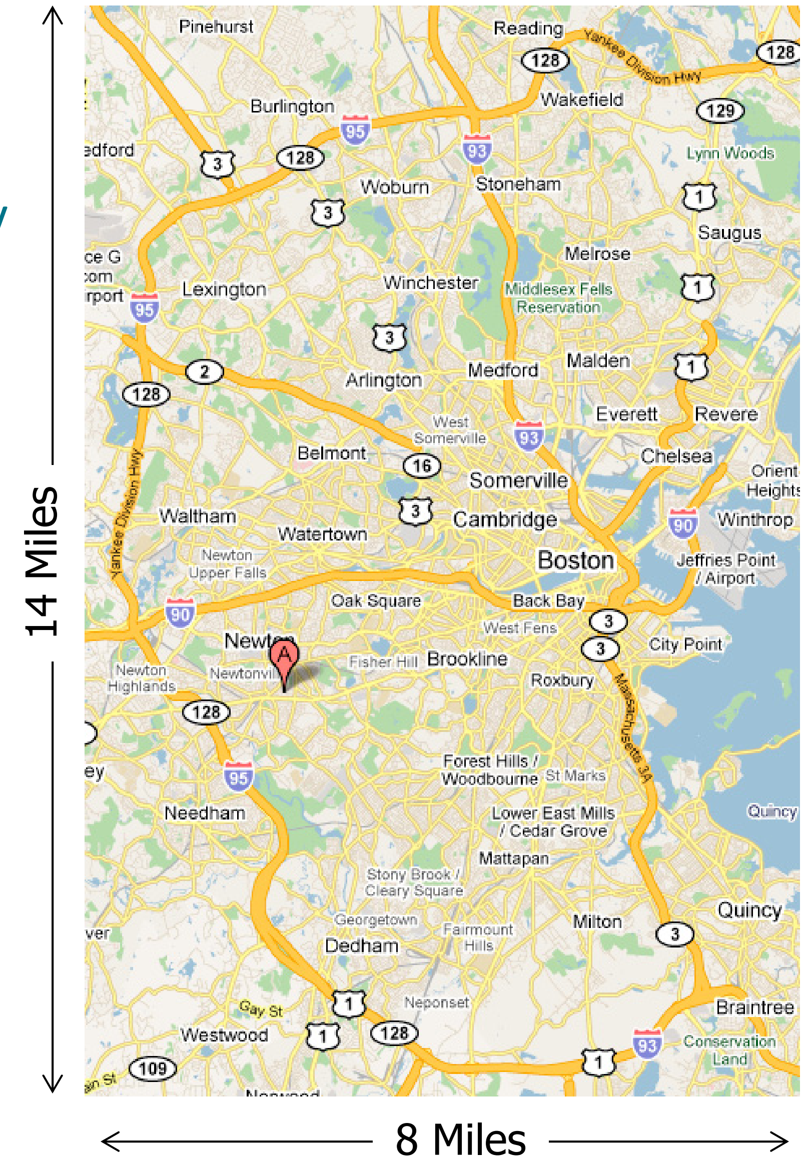
Estimating OfficeMin Costs

Estimating OfficeMin Problem

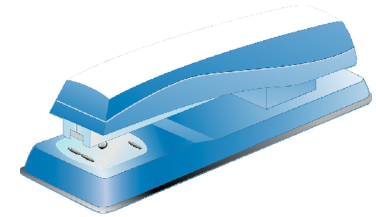


- What do we know?
  - The delivery region is about 8 by 14 miles.
  - You expect ~ 100 customer orders per day
  - Each customer orders 1 to 2 pallets of product
  - Delivery vans can handle 5 pallets at most
  - It costs ~$10 per stop (to load or unload) plus ~$5 per pallet to deliver
  - It costs about 1 $/mile to drive a van

- OK, lets try plugging and chugging!



Estimating OfficeMin Problem



$$TransportCost = c_s\left[n+\frac{D}{Q_{MAX}}+\frac{1}{2}\right]+c_d\left(2\left[\frac{D}{Q_{MAX}}+\frac{1}{2}\right]d_{LineHaul}+\frac{nk_{TSP}}{\sqrt{\delta}}\right)+c_{vs}D$$

Estimated stop (load/unload and deliver) cost per day:

$$c_s\left[n+\frac{D}{Q_{MAX}}+\frac{1}{2}\right]+c_{vs}D=10\left(100+\frac{150}{5}+.5\right)+5(150)=2055\ \$/\text{day}$$

c_s = 10 \$/stop
c_d = 1 \$/mile
c_{vs} = 5 \$/pallet
n = 100
D = 150
Q_{MAX} = 5 pallets
A = 112
δ = 100/112= .89

Estimated local driving cost per day:

Let's be conservative and assume k_{TSP} = 1.15

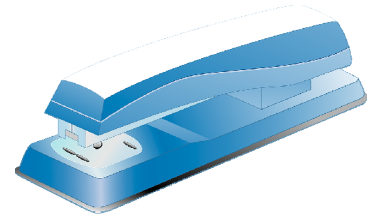
$$c_d\left(\frac{nk_{TSP}}{\sqrt{\delta}}\right)=1\left(\frac{100(k_{TSP})}{\sqrt{0.89}}\right)=105k_{TSP}\ =105(1.15)=121\ \$/\text{day}$$

Estimated linehaul driving cost per day:

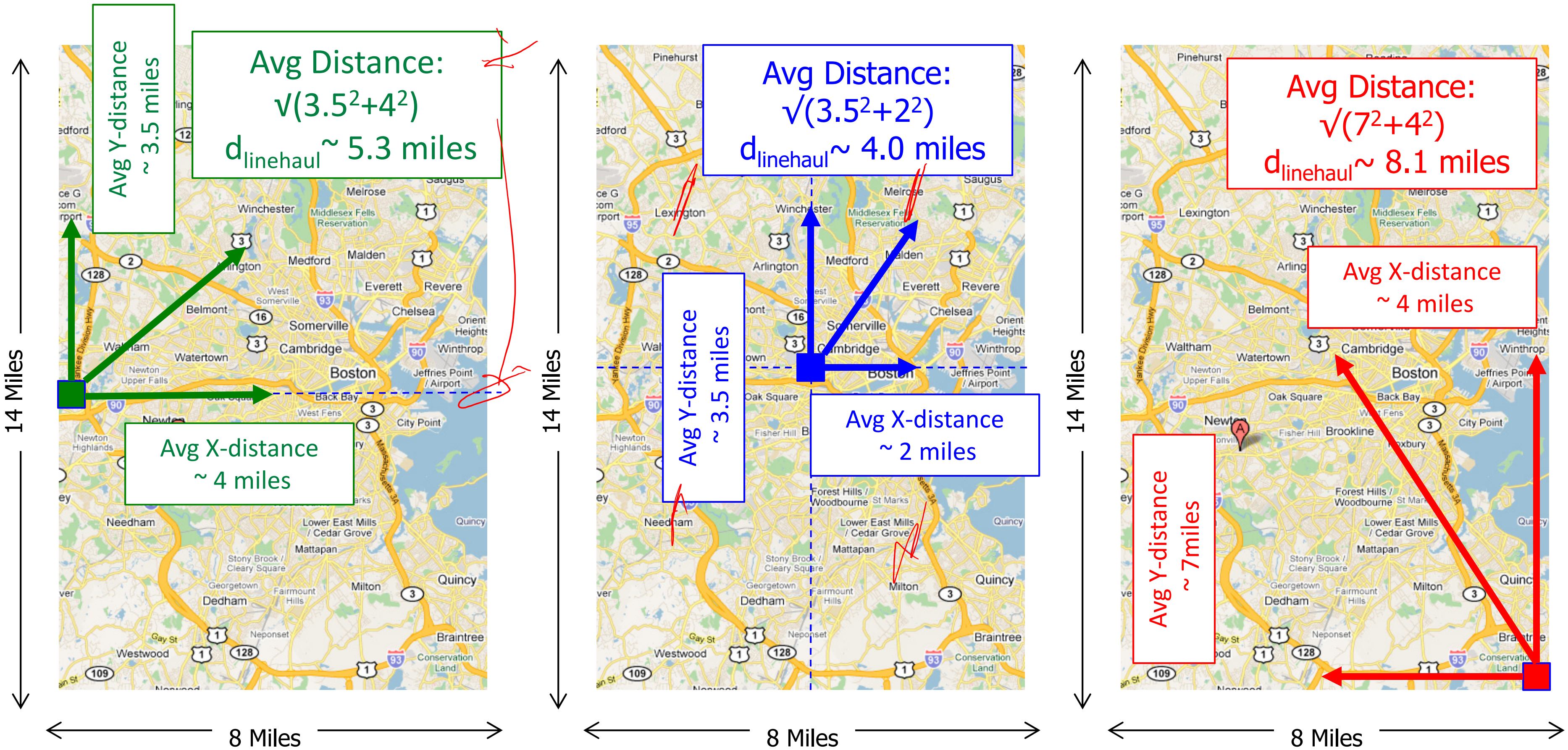
$$c_d\left(2\left[\frac{D}{Q_{MAX}}+\frac{1}{2}\right]d_{LineHaul}\right)=1\left(2\left[\frac{150}{5}+0.5\right]d_{LineHaul}\right)=61d_{LineHaul}$$

MIT Center for Transportation & Logistics

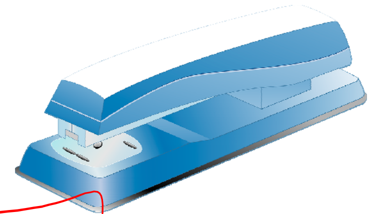
Estimating OfficeMin Problem



The expected or average linehaul distance, $d_{LineHaul}$ will depend on where the DC is located with respect to the rest of the region.



Estimating OfficeMin Problem



$$TransportCost = c_s\left[n + \frac{D}{Q_{MAX}} + \frac{1}{2}\right] + c_d\left(2\left[\frac{D}{Q_{MAX}} + \frac{1}{2}\right]d_{LineHaul} + \frac{nk_{TSP}}{\sqrt{\delta}}\right) + c_{vs}D$$

Estimated stop (load/unload and deliver) cost = **2055** \$/day

Estimated local driving cost = **121** \$/day

Estimated linehaul driving cost = **323** \$/day

$$c_d\left(2\left[\frac{D}{Q_{MAX}} + \frac{1}{2}\right]d_{LineHaul}\right) = 1\left(2\left[\frac{150}{5} + 0.5\right]d_{LineHaul}\right) = 61d_{LineHaul} = 61(5.3) = 323 \text{ \$/day}$$

Estimated total transportation cost per day ~ **2500** \$/day

So what? What can I do with this?

- Is this even in the ballpark of reason to continue?
- Where should we focus efforts?
- How sensitive is the estimate to my assumptions?

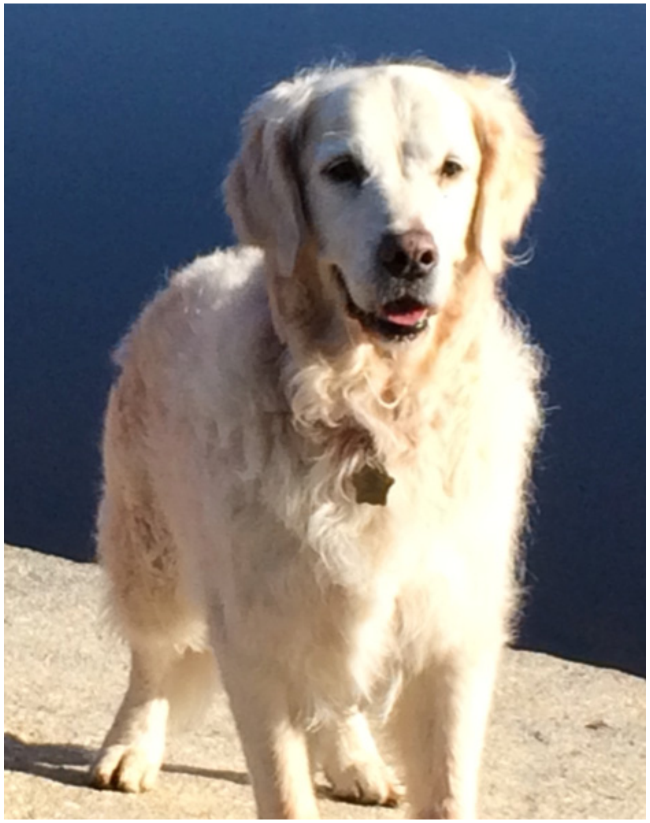
MIT Center for Transportation & Logistics

Key Take Aways

Key Take Aways

- Approximations are good "first steps"
  - Require minimal data
  - Allow for fast sensitivity analysis
  - Enables quick scoping of the solution space
- Quick estimation –
  - Break into pieces, Estimate independently, Recombine
- Think of Goldilocks:
  - Too small (pass), Too large (do it), Just right (look into)
- Optimal methods usually require tremendous amounts of detailed information
- Before deciding to spend the time and energy to find an optimal solution, it is helpful to see if it is worth it.

Questions, Comments, Suggestions?
Use the Discussion Forum!



"Dexter – continuously approximating to his food bowl"
Yankee Golden Retriever Rescued Dog (www.ygrr.org)

MIT Center for Transportation & Logistics

caplice@mit.edu