



MONASH
University

**MONASH
BUSINESS
SCHOOL**

Business Analytics

Professor Di Cook
Econometrics and Business Analytics

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of commerce marketing. Our fields of expertise include: marketing strategy and optimisation, customer tracking and on-site analytics, predictive analytics and econometrics, data warehousing and big data systems, marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY
© Krzysztof Zawadzki

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Matrix algebra
- ★ Parallel databases and parallel processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with Scala/Spark

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

the sexiest job of the 21st century, requires a mix of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business.

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of commerce marketing. Our fields of expertise include: marketing strategy and optimisation, customer tracking and on-site analytics, predictive analytics and econometrics, data warehousing and big data systems, marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY
© Krzysztof Zawadzki

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Hadoop, NoSQL, Tableau

machine learning, statistical modelling, experimental design, bayesian inference, supervised and unsupervised learning, optimisation

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of commerce marketing. Our fields of expertise include: marketing strategy and optimisation, customer tracking and on-site analytics, predictive analytics and econometrics, data warehousing and big data systems, marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY
© Krzysztof Zawadzki

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influential without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

Scripting languages for data analysis and statistical modeling like R

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of commerce marketing. Our fields of expertise include: marketing strategy and optimisation, customer tracking and on-site analytics, predictive analytics and econometrics, data warehousing and big data systems, marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY
© Krzysztof Zawadzki

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influential without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

COMMUNICATION & VISUALIZATION

- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

Curious about data
Hacker mindset
Problem solver

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of commerce marketing. Our fields of expertise include: marketing strategy and optimisation, customer tracking and on-site analytics, predictive analytics and econometrics, data warehousing and big data systems, marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY
© Krzysztof Zawadzki

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Matrix/algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with cloud like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

R packages like ggplot2
Story-telling skills

What is a business analyst?

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experimental design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Linear algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience withaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influences without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Highcharts, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of commerce marketing. Our fields of expertise include: marketing strategy and optimisation, customer tracking and on-site analytics, predictive analytics and econometrics, data warehousing and big data systems, marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY
© Krzysztof Zawadzki

Top 10 jobs 2017

- 1** *Actuary* **2015**
- 2** *Audiologist*
- 3** *Mathematician*
- 4** *Statistician*
- 5** Biomedical Engineer
- 6** **Data Scientist**
- 7** Dental Hygienist
- 8** Software Engineer
- 9** Occupational Therapist
- 10** Computer Systems Analyst

2015 Career Best: <http://www.careercast.com>

Top 10 jobs 2017

1. **Data Scientist (\$USD128,240, proj. growth 16%)**
 2. Statistician (\$80,295, 37%)
 3. Information Security Analyst (\$89,280, 18%)
 4. Audiologist
 5. Diagnostic Medical Sonographer
 6. Mathematician
 7. Software Engineer
 8. Computer Systems Analyst
 9. Speech Pathologist
 10. Actuary
- 2016

<http://www.careercast.com/jobs-rated/jobs-rated-report-2016-ranking-200-jobs>

Top 10 jobs 2017

1. Statistician (\$80,110, proj. growth 34%)
2. Medical Services Manager (\$94,500, 17%)
3. Operations Research Analyst (\$79,200, 30%)
4. Information Security Analyst (\$90,120, 18%)
5. **Data Scientist (\$111,267, 16%)**
6. University Professor (\$72,416, 15%)
7. Mathematician (\$111,298, 22%)
8. Software Engineer (\$100,690, 17%)
9. Occupational Therapist (\$81,910, 29%)
10. Speech Pathologist (\$73,250, 27%)

<http://www.careercast.com/jobs-rated/best-jobs-2017>

Projects

- Math gap
- Climate change
- Pedestrian sensors
- Curating art exhibits

Numbers point to maths 'gap'

May 2, 2011

☆ Read later

Caroline Milburn

[Tweet](#) [Share 0](#) [G+ Share 0](#) [Share](#) [Pin it](#) [submit](#)

[Email article](#) [Print](#) [Reprints & permissions](#)



Australia is one of the few countries with a maths gap in favour of boys.

GIRLS are performing at much lower levels in maths than boys — and the gap is widening as students progress through school.

A study of NAPLAN numeracy test results for students in years 3, 5, 7 and 9, to be presented

“Girls are performing at much lower levels in maths than boys - and the gap is widening as students progress through school.”
SMH, 2011

Gender Gap

- OECD PISA survey “the world’s global metric for quality, equity and efficiency in school education”.
- Workforce readiness of 15-year old students
- 500,000 students were tested across 65 countries and 18,000 schools
- Math, reading and science
- Data available from <http://www.oecd.org/pisa/data/>
- Are boys better than girls at math?

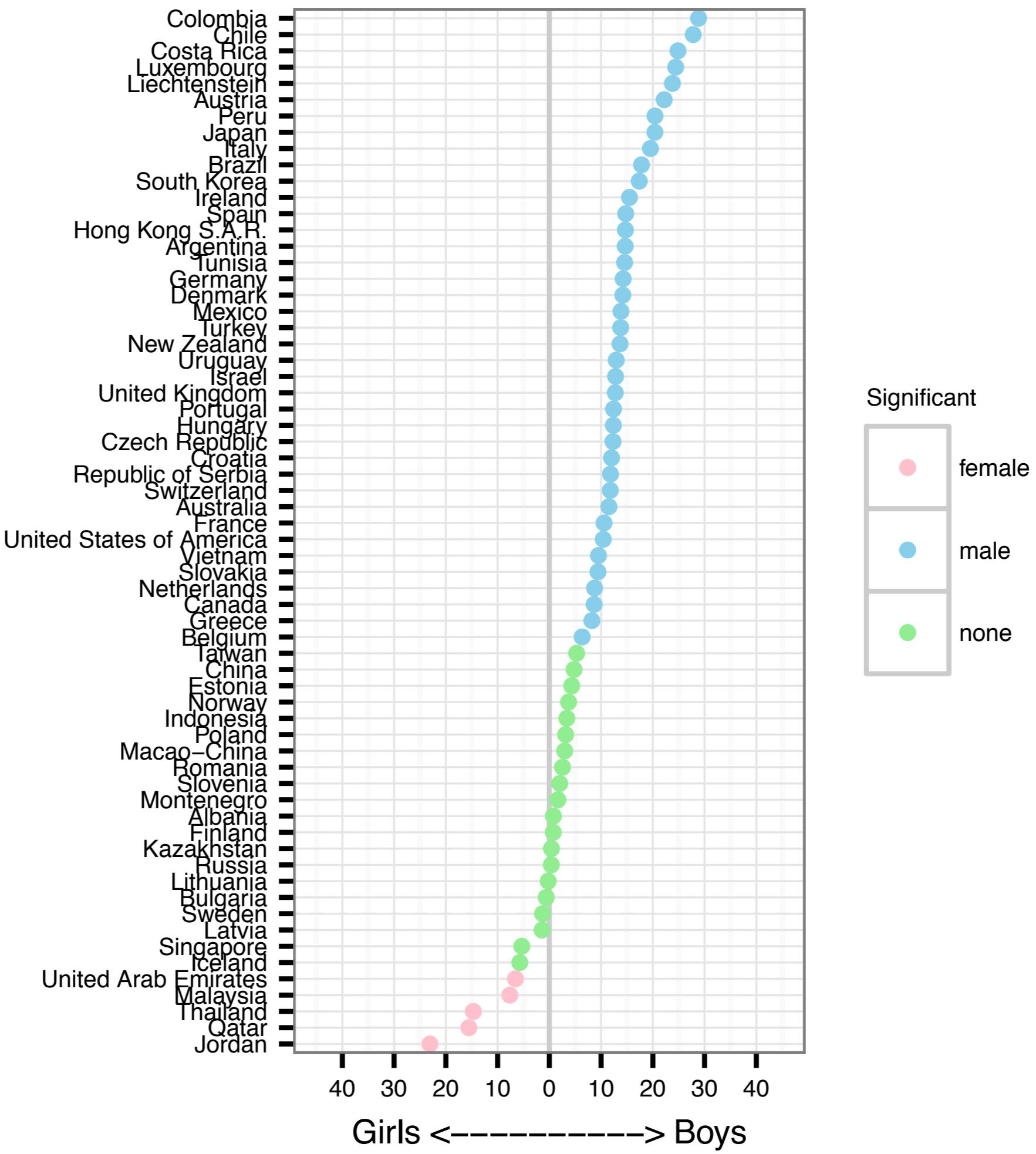
Calculations

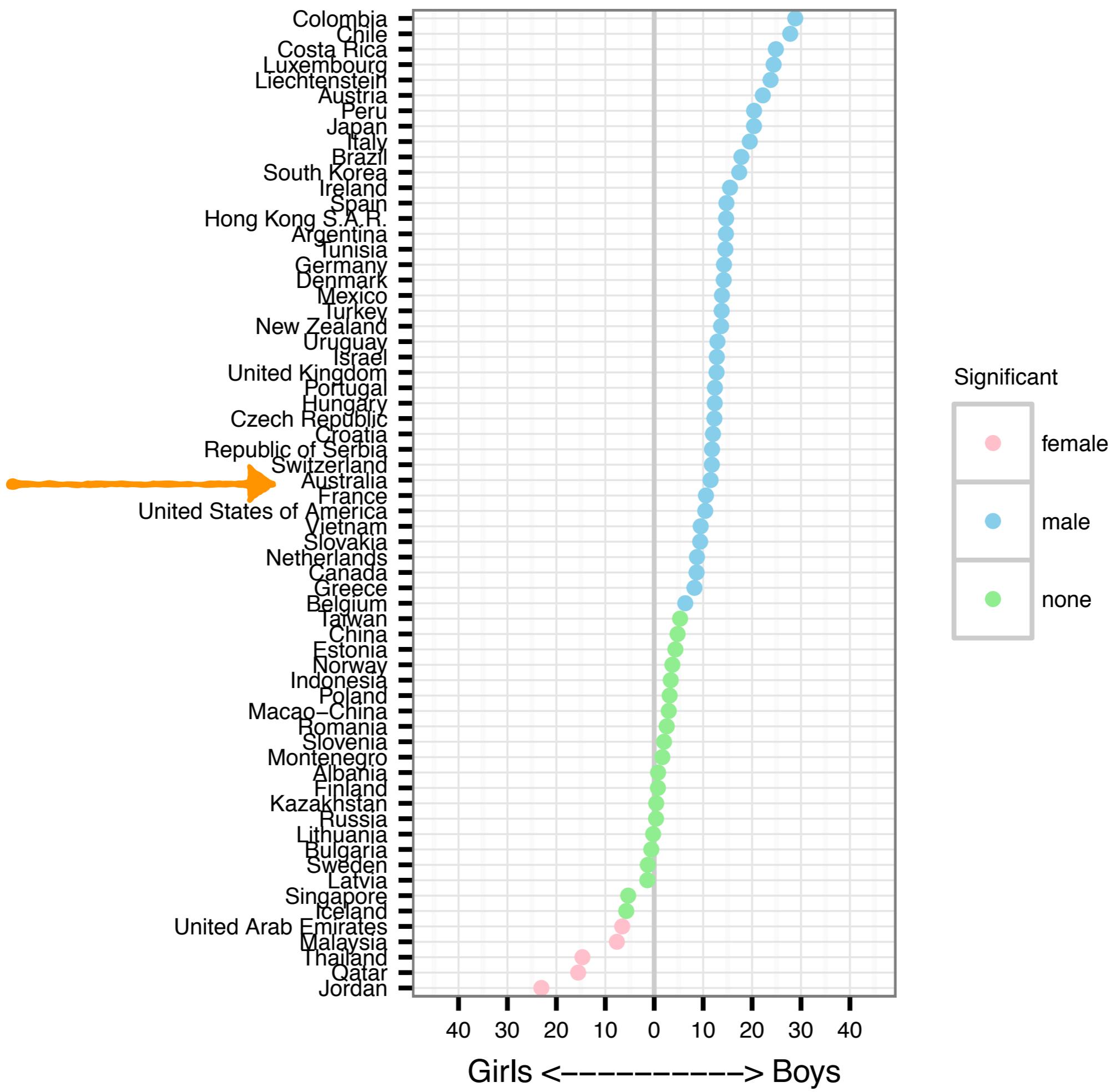
- Difference the average math scores, using a weighted mean, because the tests are selected using a sampling framework. The weights indicate how representative they are of the cohort in their country

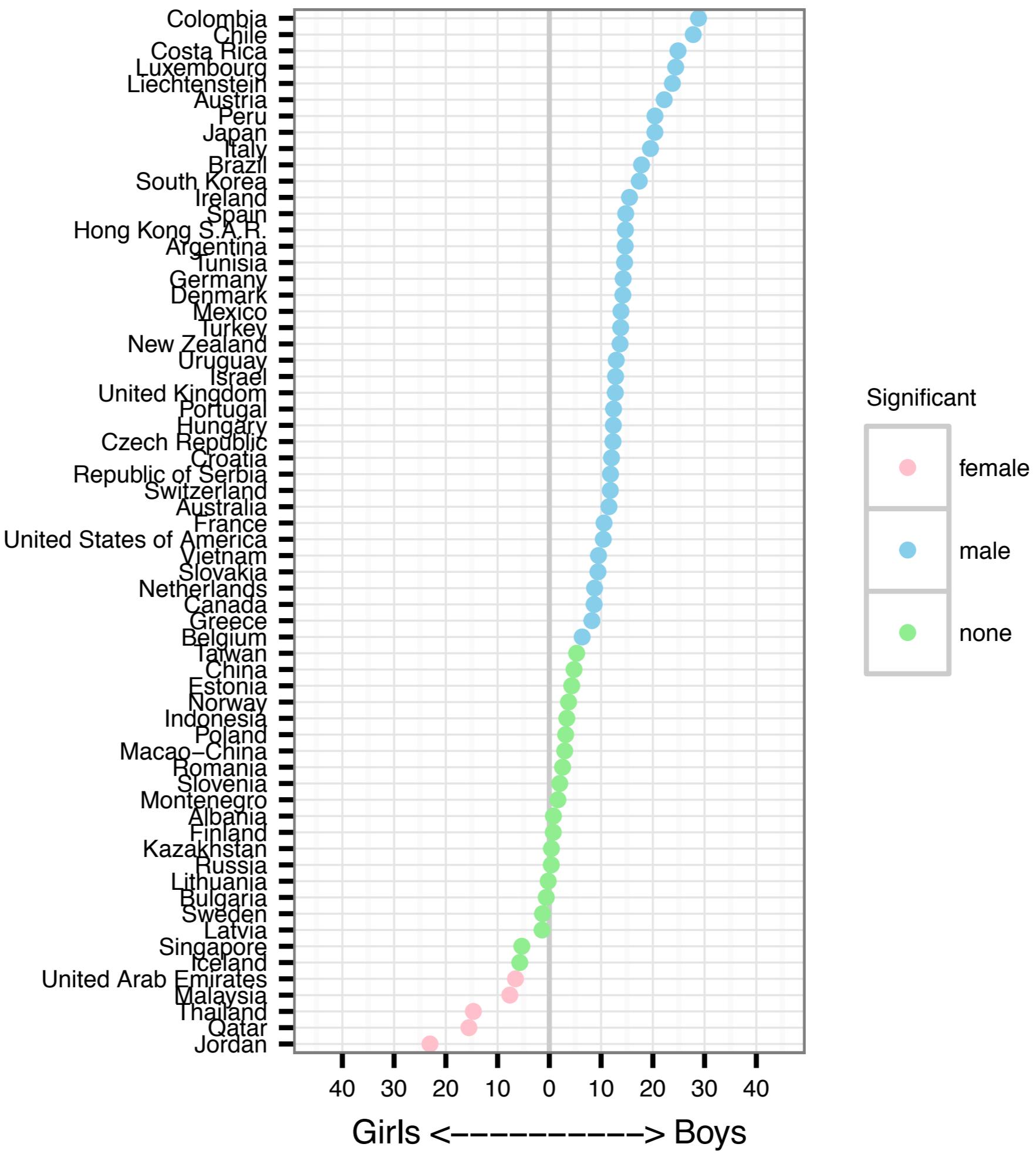
$$\frac{1}{n} \sum_{i=1}^n w_i x_i$$

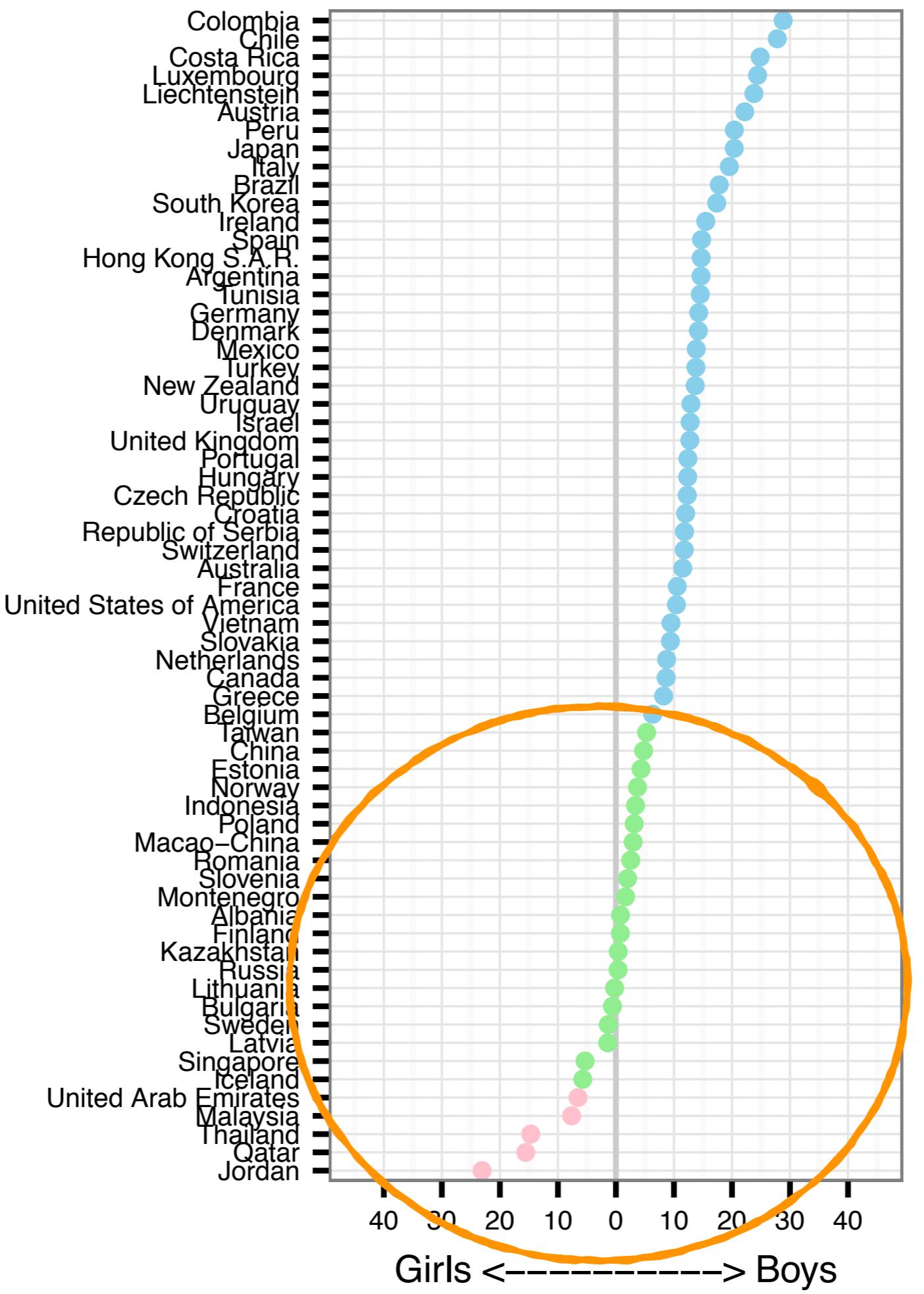
$$\bar{x}_{boys} - \bar{x}_{girls}$$

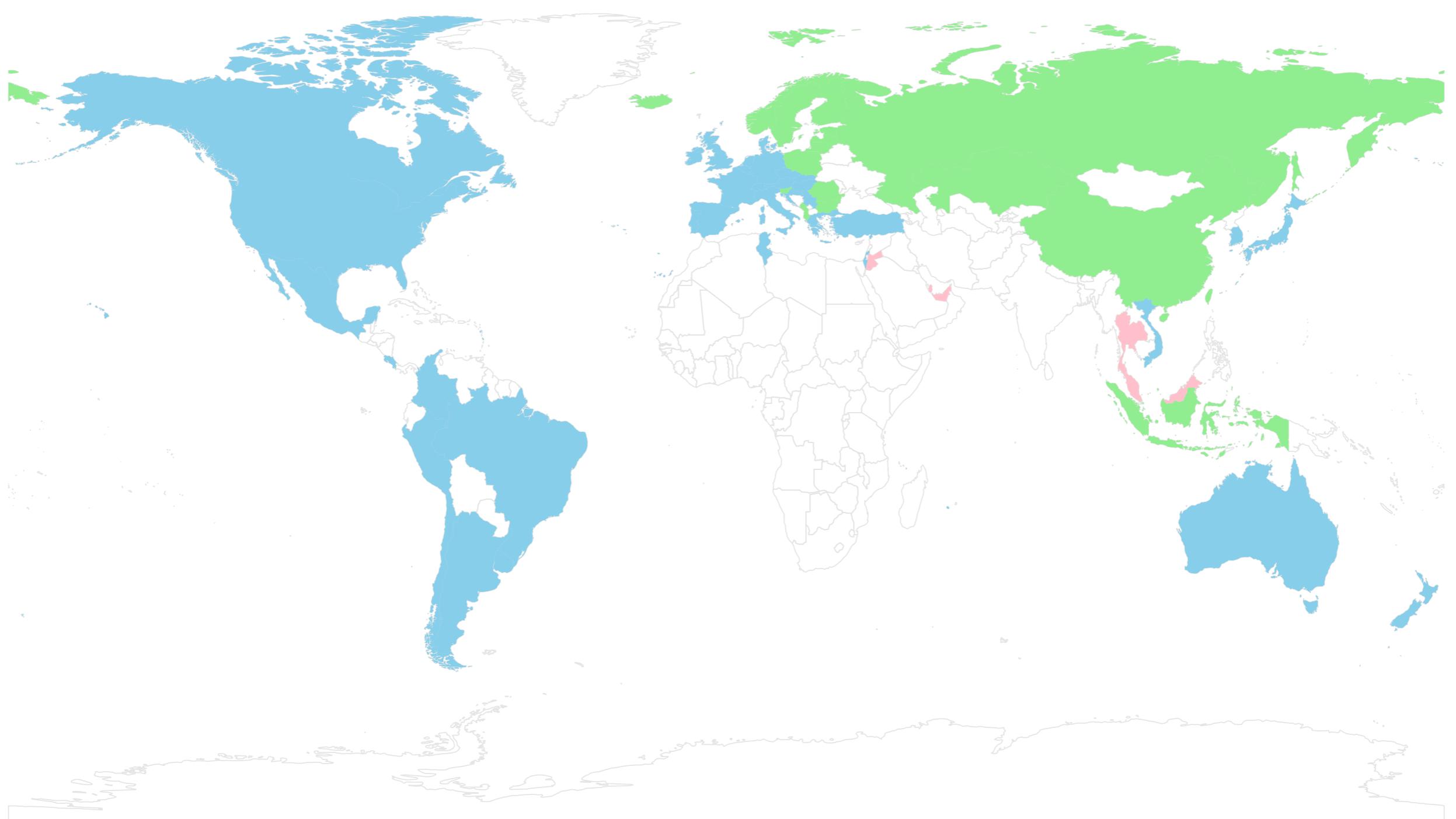
- Calculate for each country, and statistically test the difference











UN Climate Conference Weather Climate Change Whale Watch Animals Conservation Energy Sm

You are here: Home > Environment >

CSIRO team's study erodes credibility of key soil carbon model

November 1, 2015

 Read later



Peter Hannam

Environment Editor, The Sydney Morning Herald

[View more articles from Peter Hannam](#)

 Follow Peter on Twitter  Follow Peter on Google+  Email Peter

 Tweet  Share 459  Share 11  Share 1  Pin it submit

 Email article  Print  Reprints & permissions



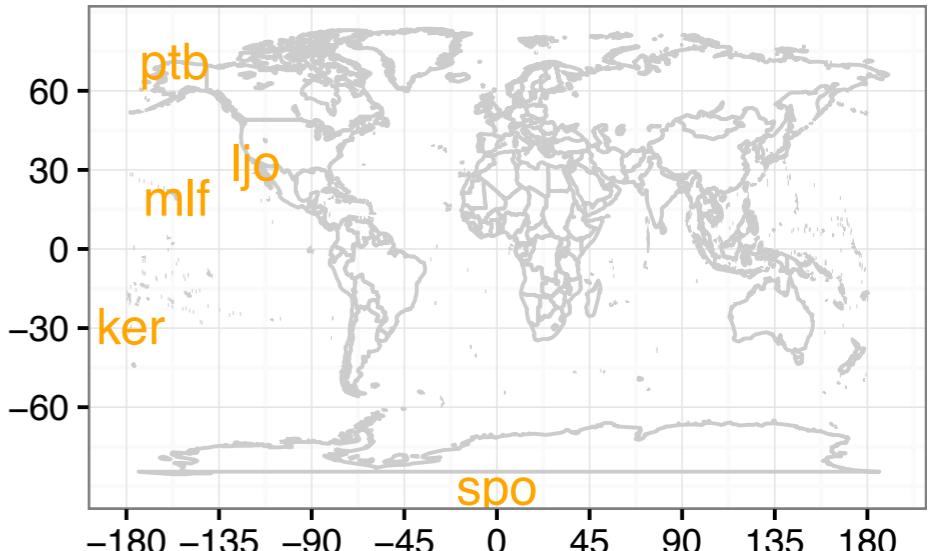
Sydney Barrister Peter King argues a royal commission into banks' treatment of farmers is needed. Photo: Jessica Shapiro

Australia's method of measuring how much carbon is being stored in its soil is flawed, undermining the credibility of government programs to pay farmers to sequester the climate change inducing element, a new study by CSIRO researchers has found.

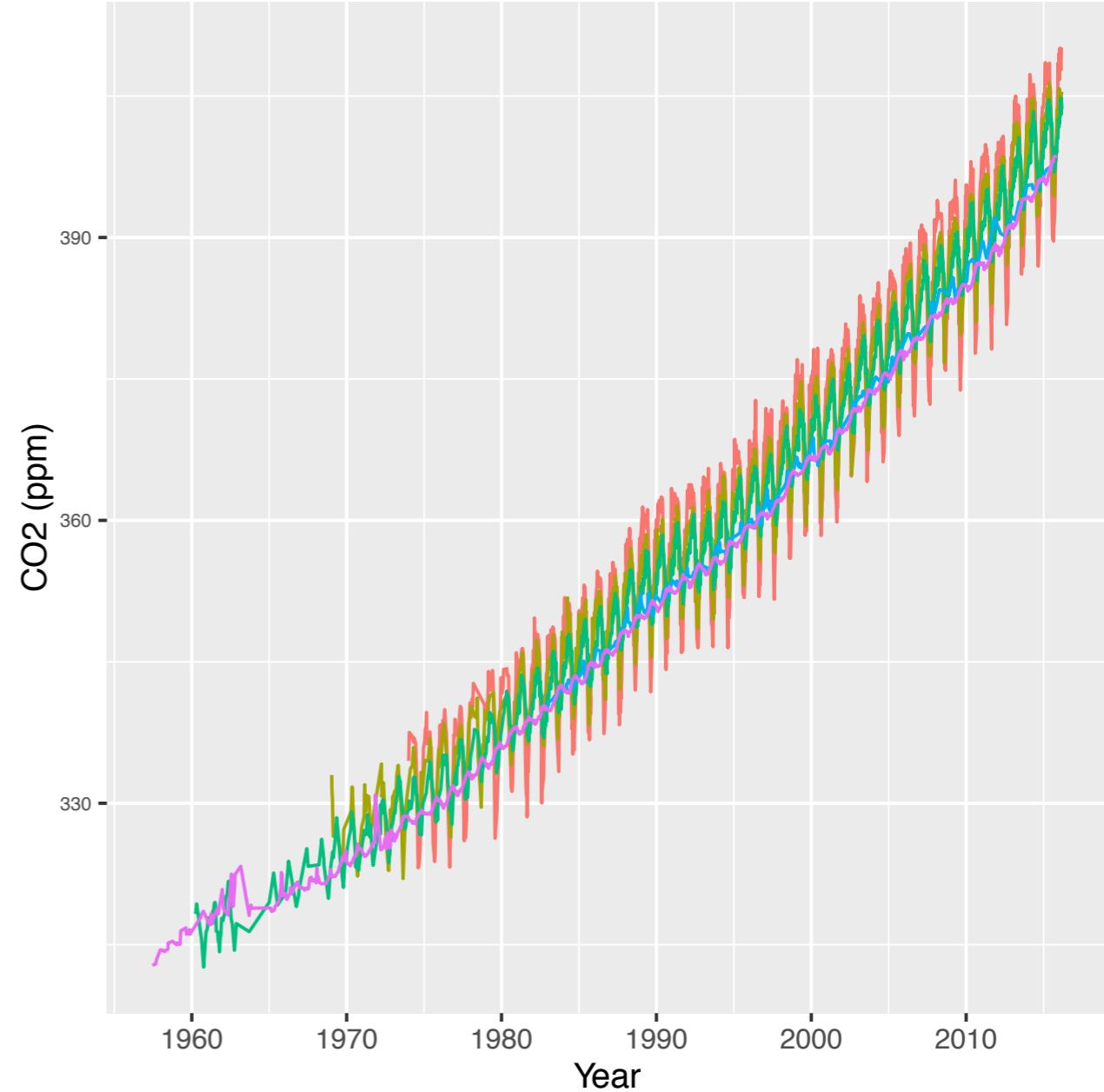
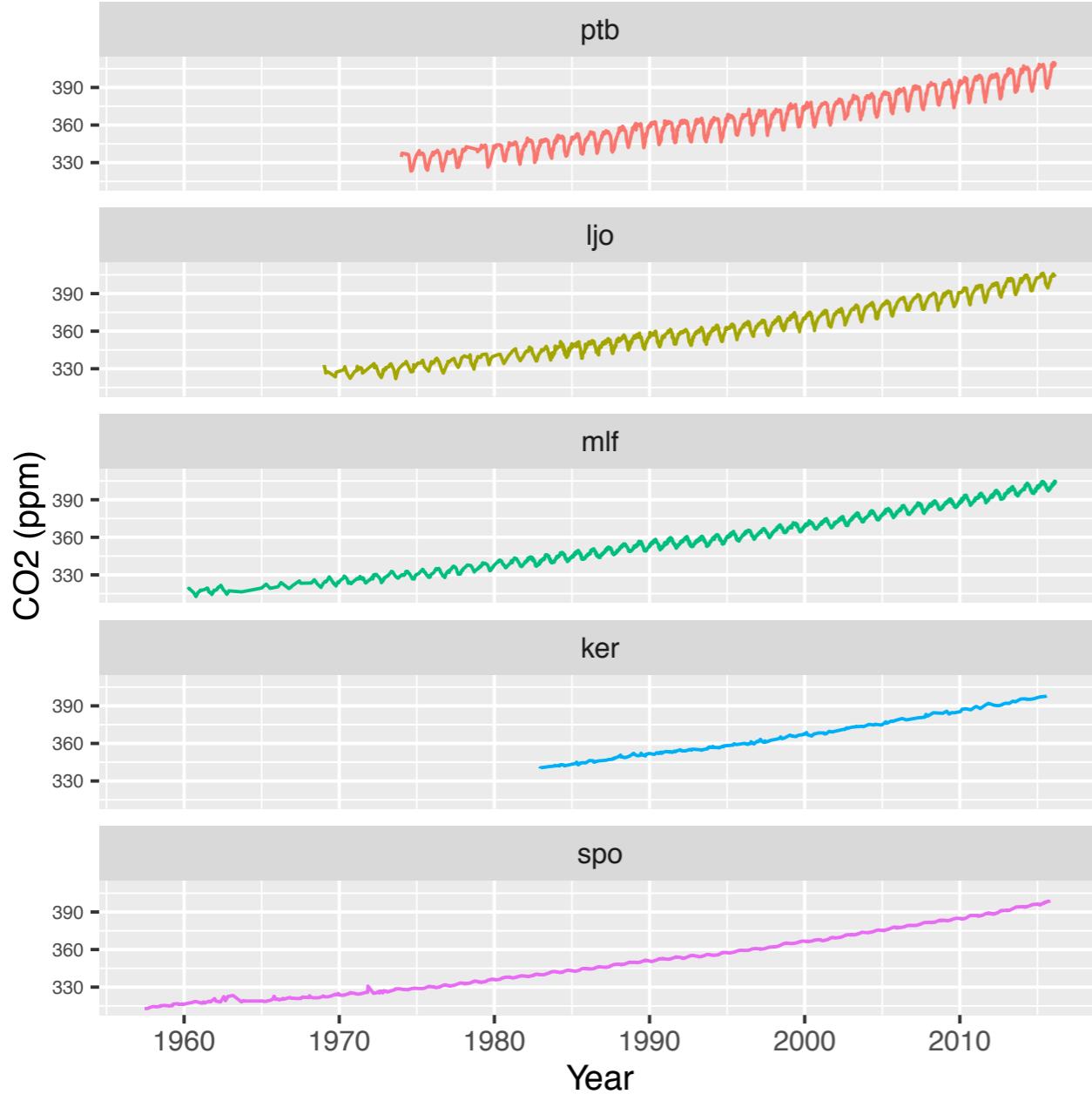
"Australia's method of measuring how much carbon is being stored in its soil is flawed, undermining the credibility of government programs to pay farmers to SEQUESTER the climate change inducing element."

CO₂

- Data is collected at a number of locations world wide. See Scripps Inst. of Oceanography (http://scrippSCO2.ucsd.edu/data/atmospheric_co2)
- Pull the data from this site and plot it
- Recordings from South Pole (SPO), Kermadec Islands (KER), Mauna Loa Hawaii (MLF), La Jolla Pier, California (LJO), Point Barrow, Alaska (PTB).



CO₂

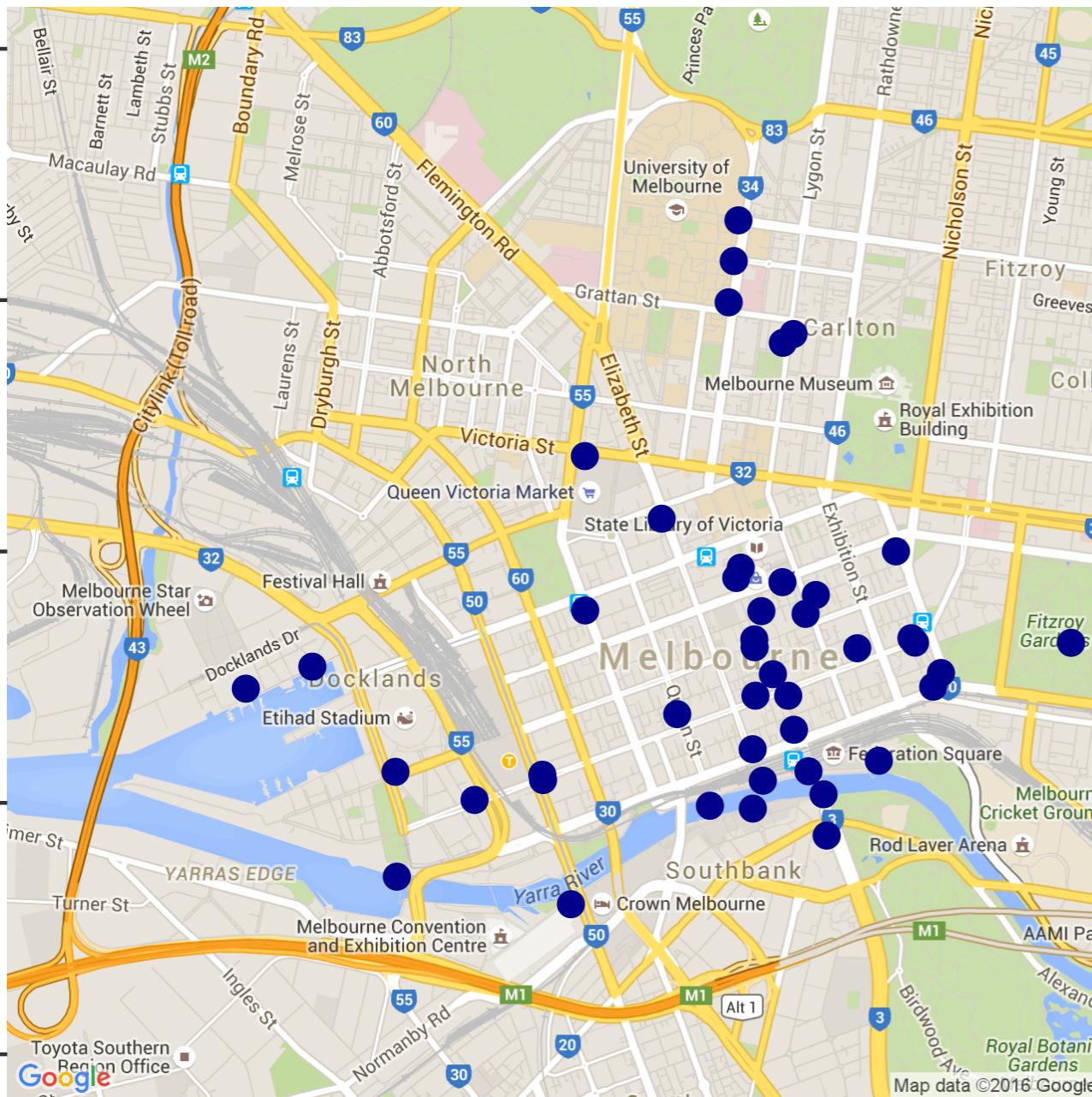


CO₂

- CO₂ is increasing, and it looks like it is exponential increase. I really expected that the concentration would have flattened out with all of the efforts to reduce carbon emissions.
- The same trend is seen at every location - REALLY? Need some physics to understand this.
- Some stations show seasonal pattern - actually the more north the more seasonality - WHY?

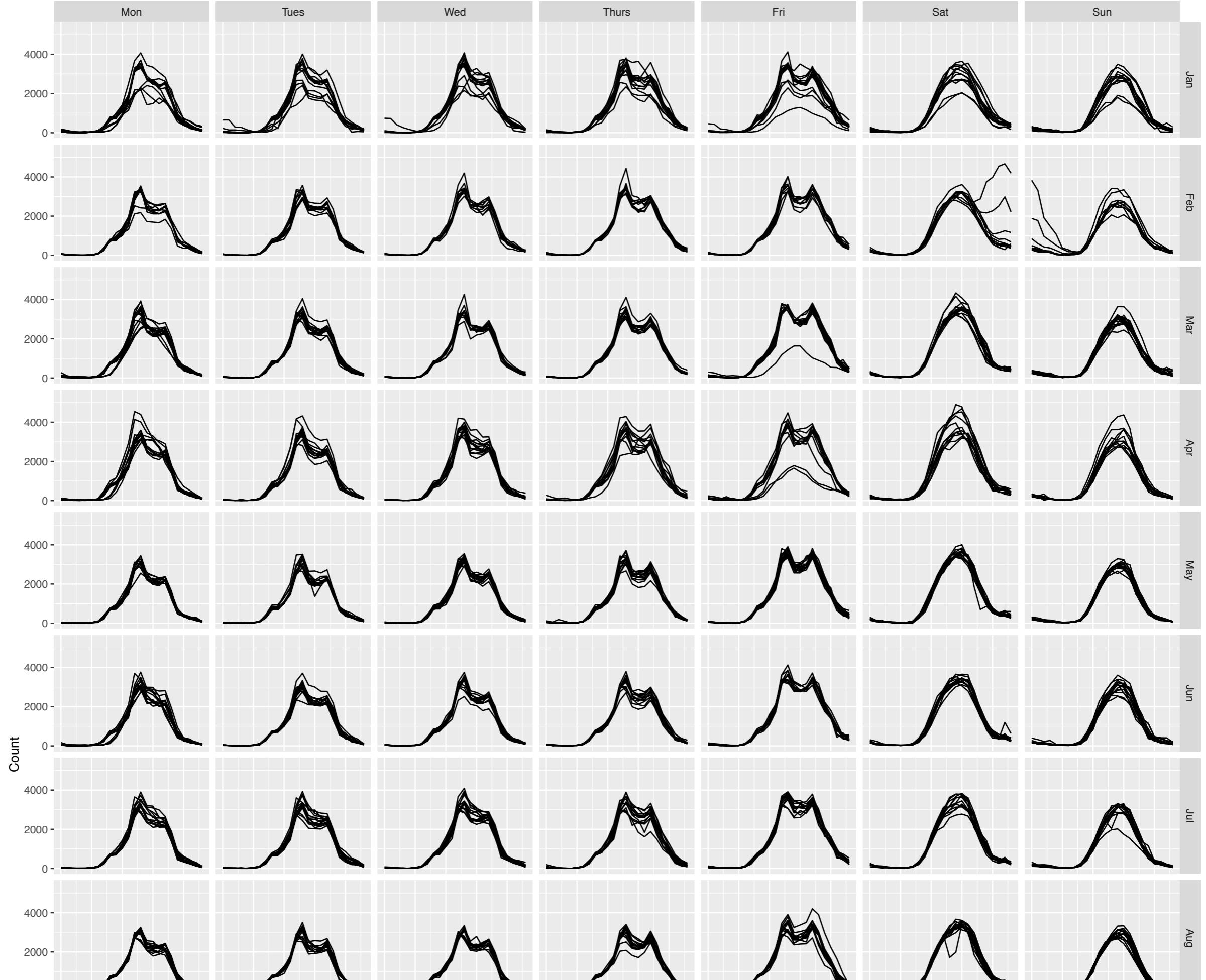
Pedestrian sensors

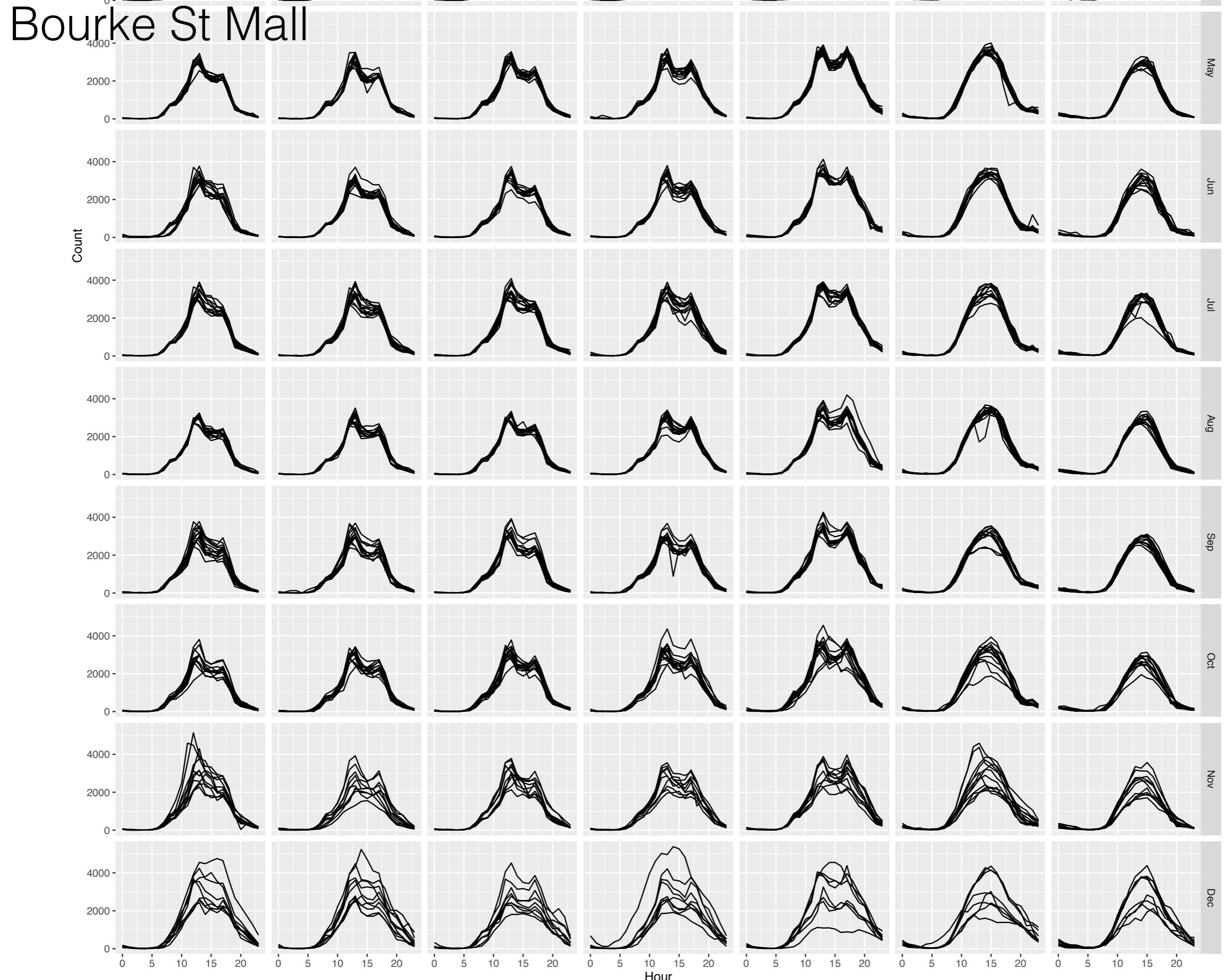
Sensor locations



- Measuring passers-by
- Counts reported every hour
- Some recording since 2009

Bourke St Mall





Airline traffic

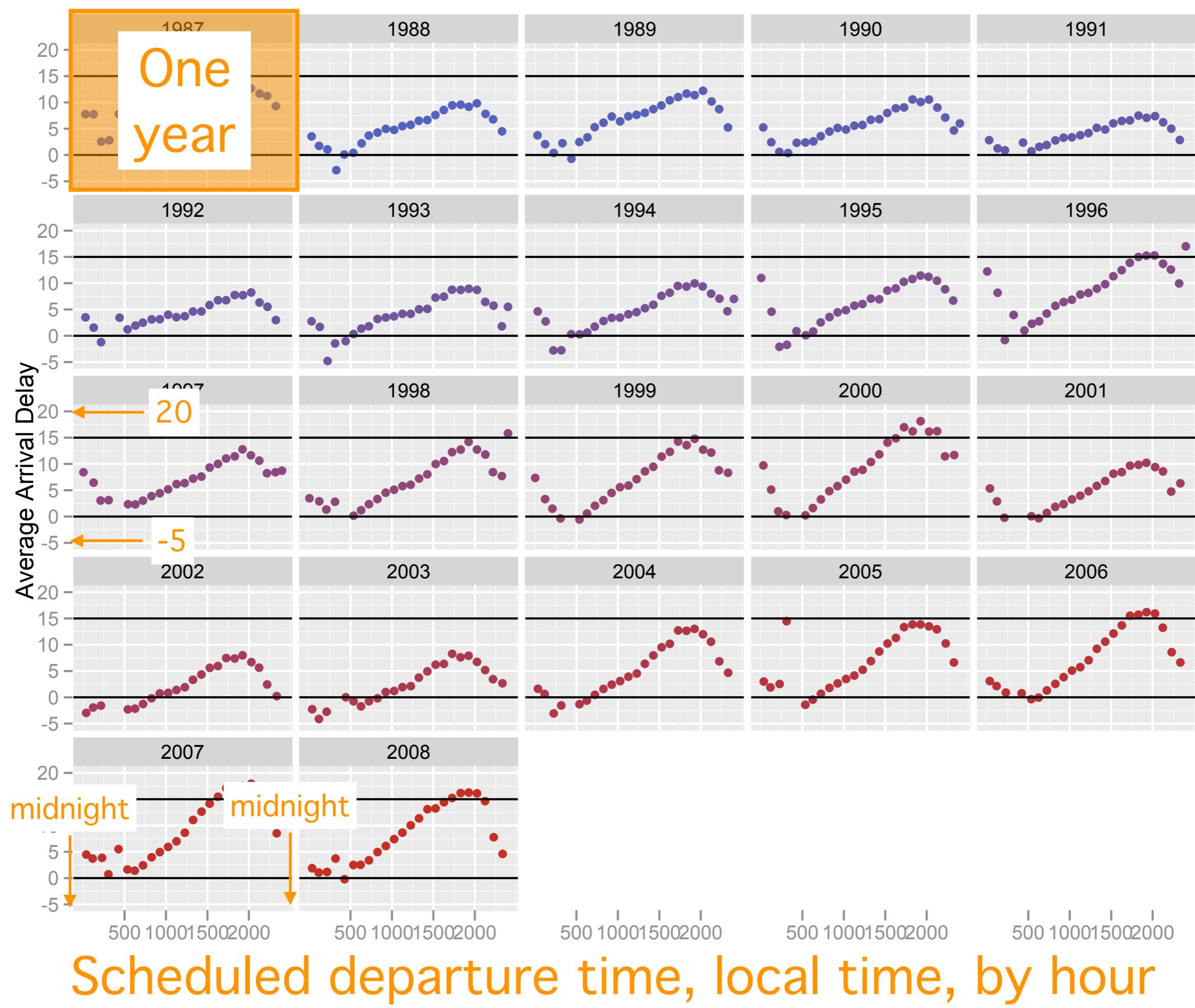
- ~15,000 flights a day
- April 1986 - present
- RITA - Research and Innovative Technology Administration (flight information, arrival delay, airline, plane id, ...)
- On time performance database - <http://www.transtats.bts.gov/> You can download this yourself!

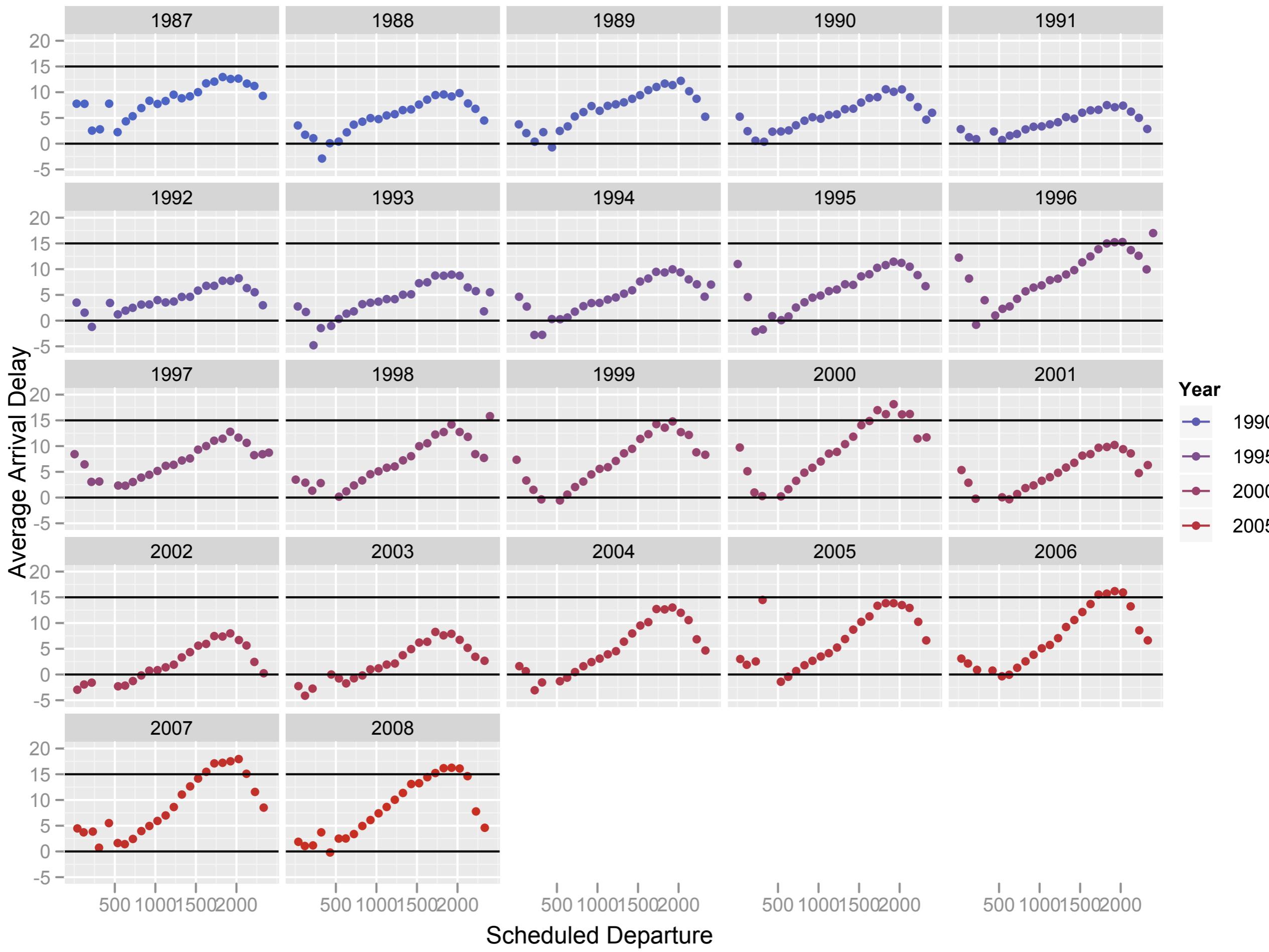
Airline traffic

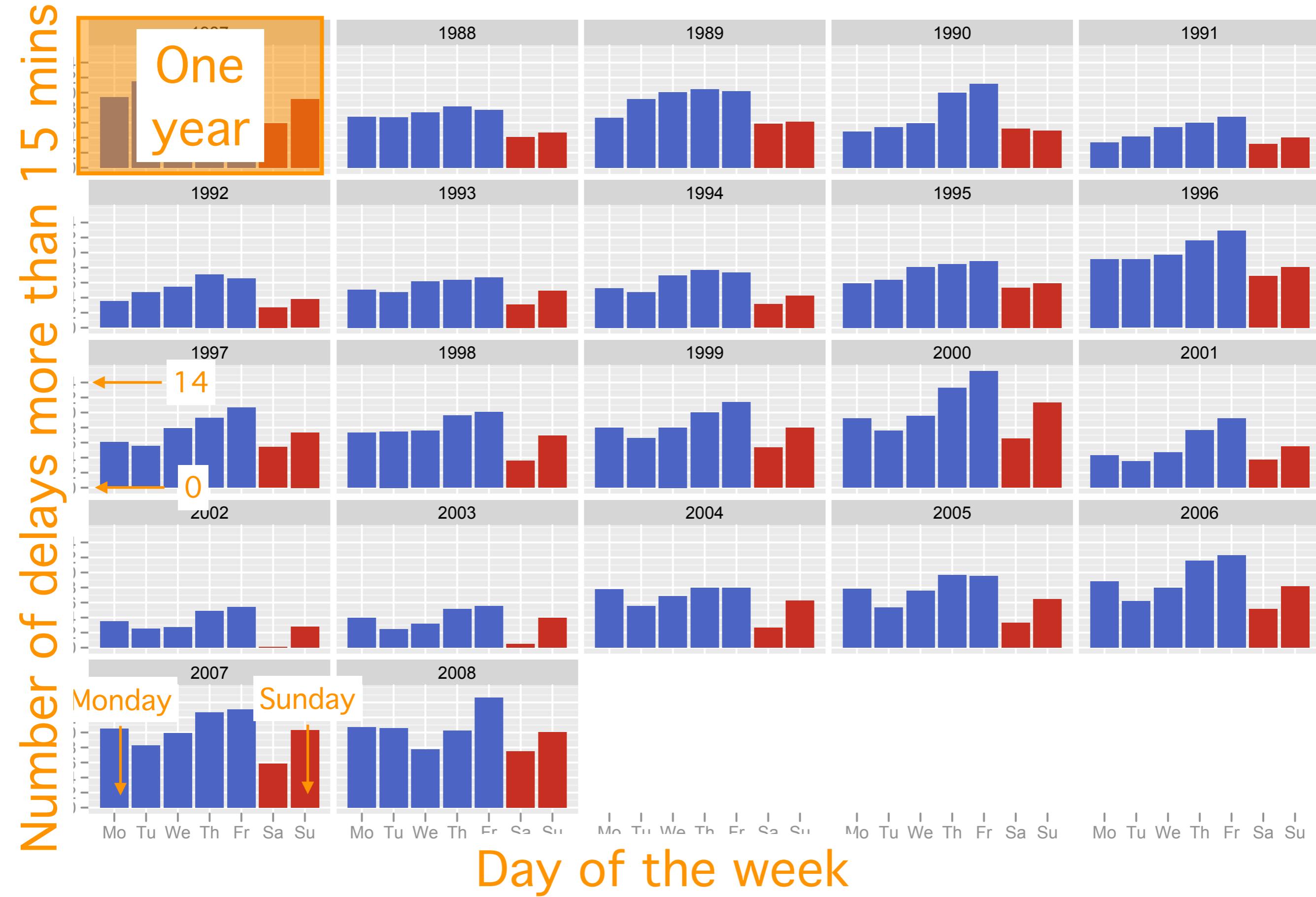
- ~15,000 flights a day
- April 1986 - present
- RITA - Research and Innovative Technology Administration (flight information, arrival delay, airline, plane id, ...)
- On time performance database - <http://www.transtats.bts.gov/> You can download this yourself!

You can download this yourself!

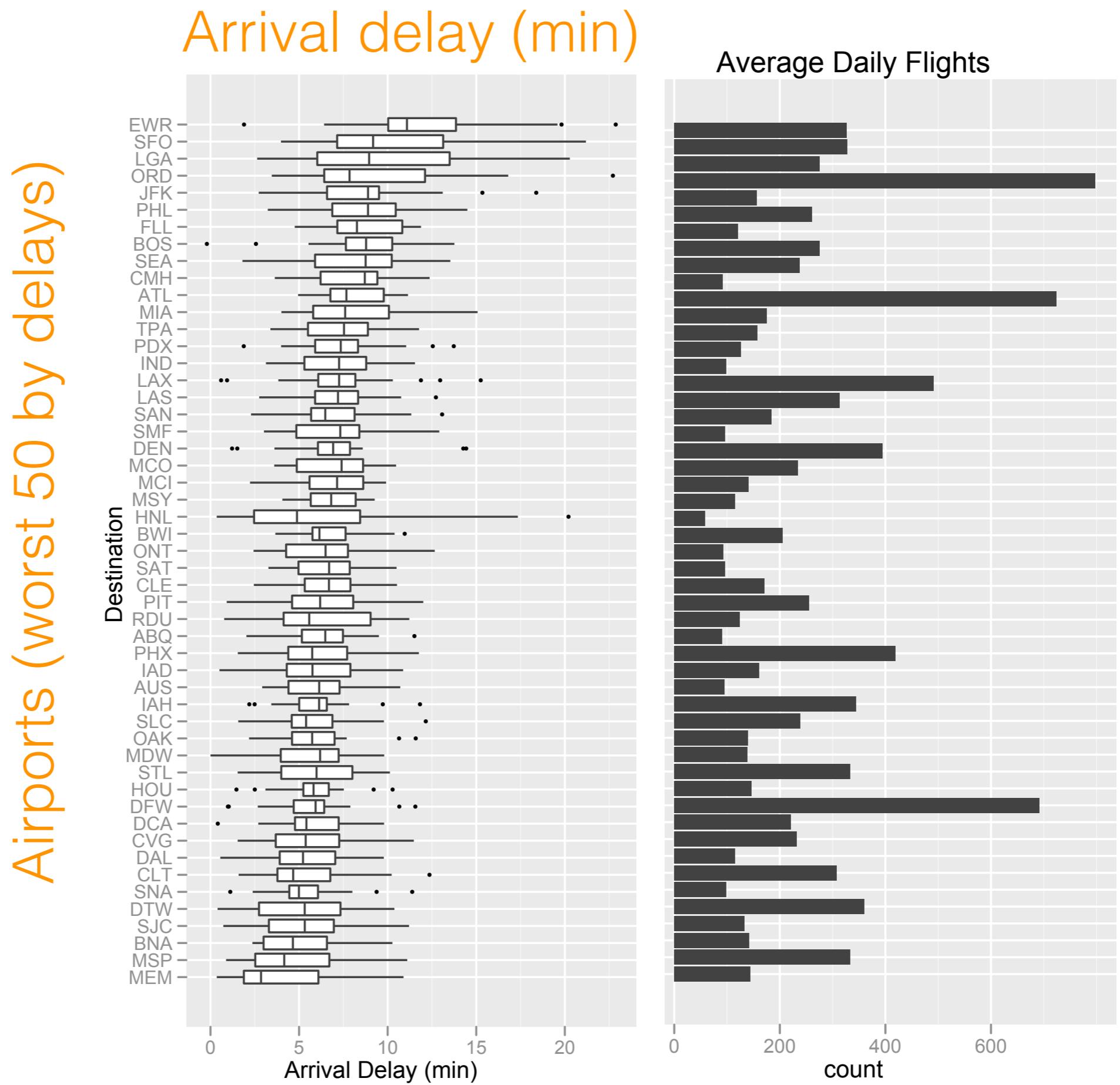
Average arrival delay, minutes



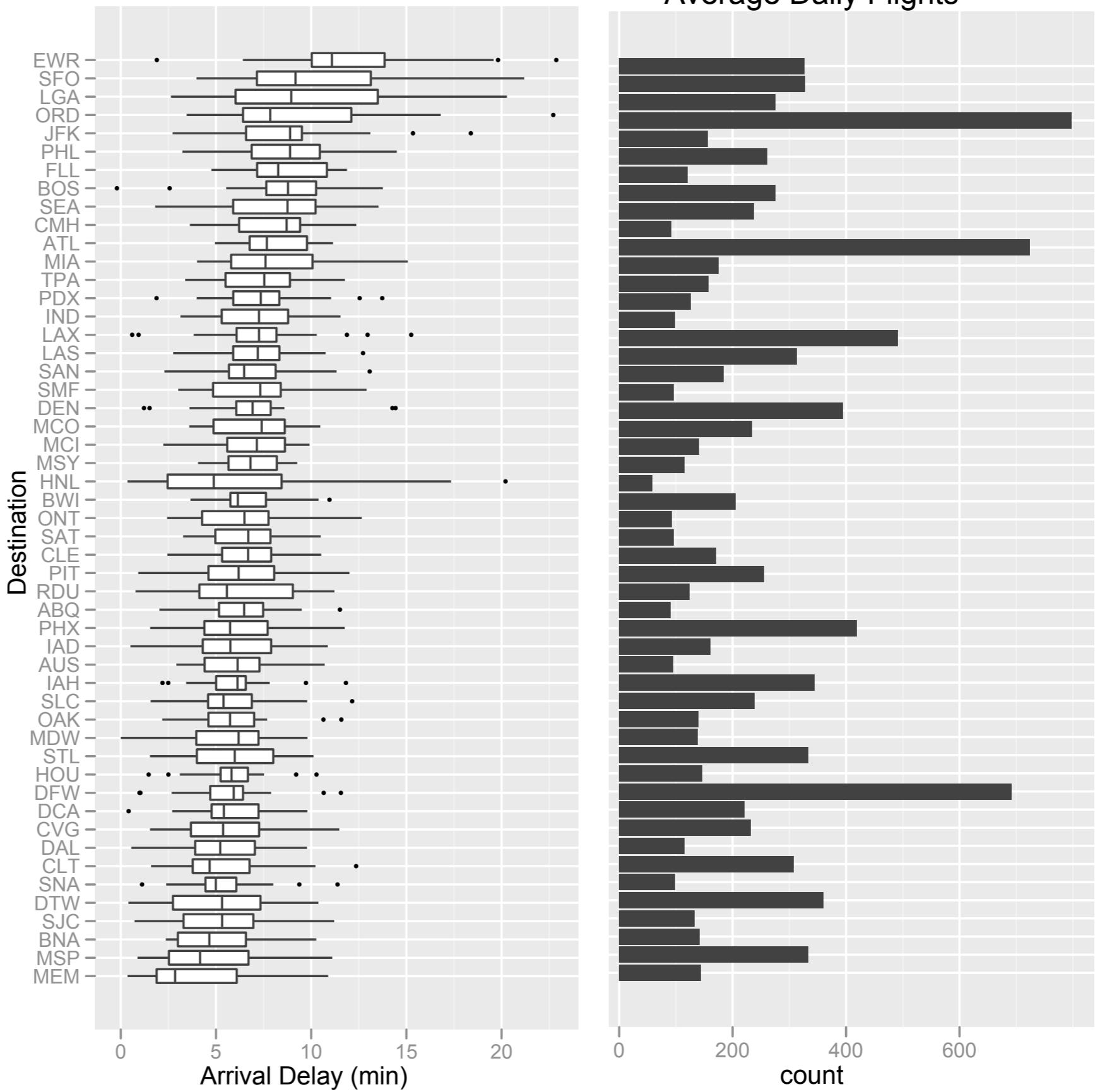


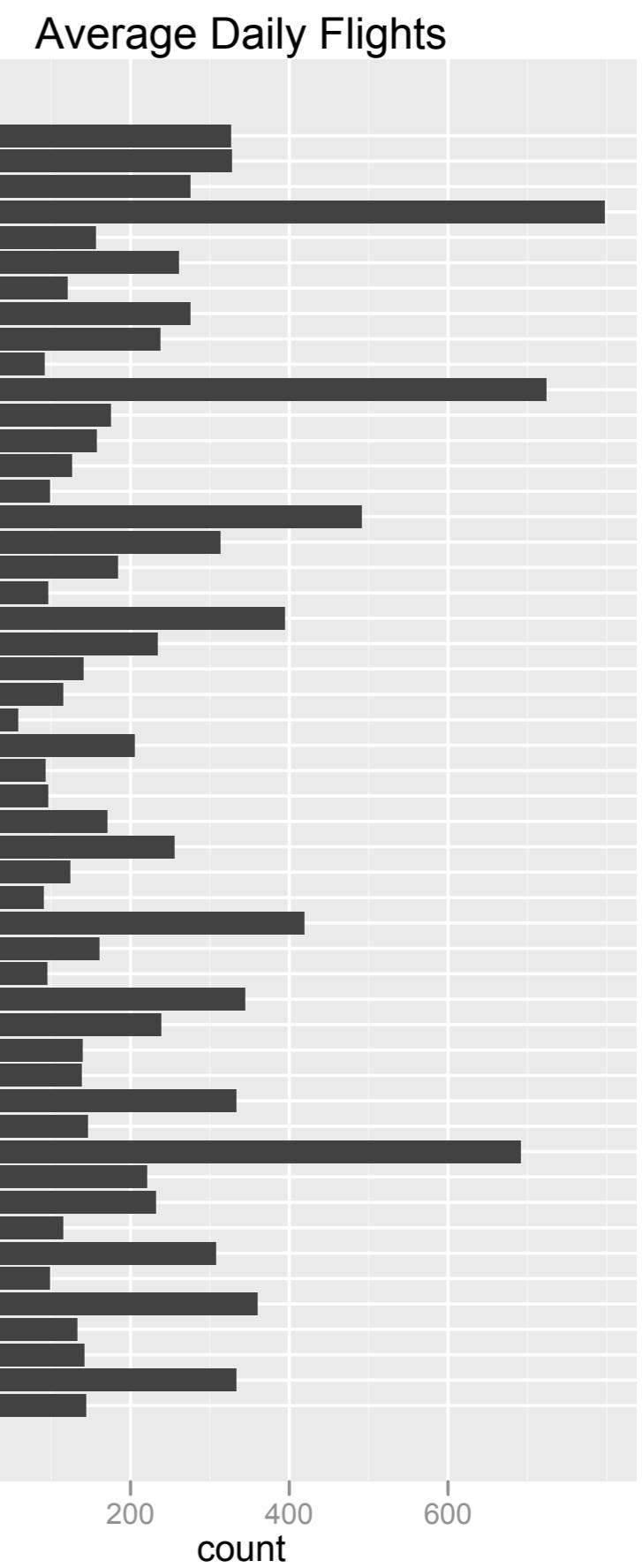
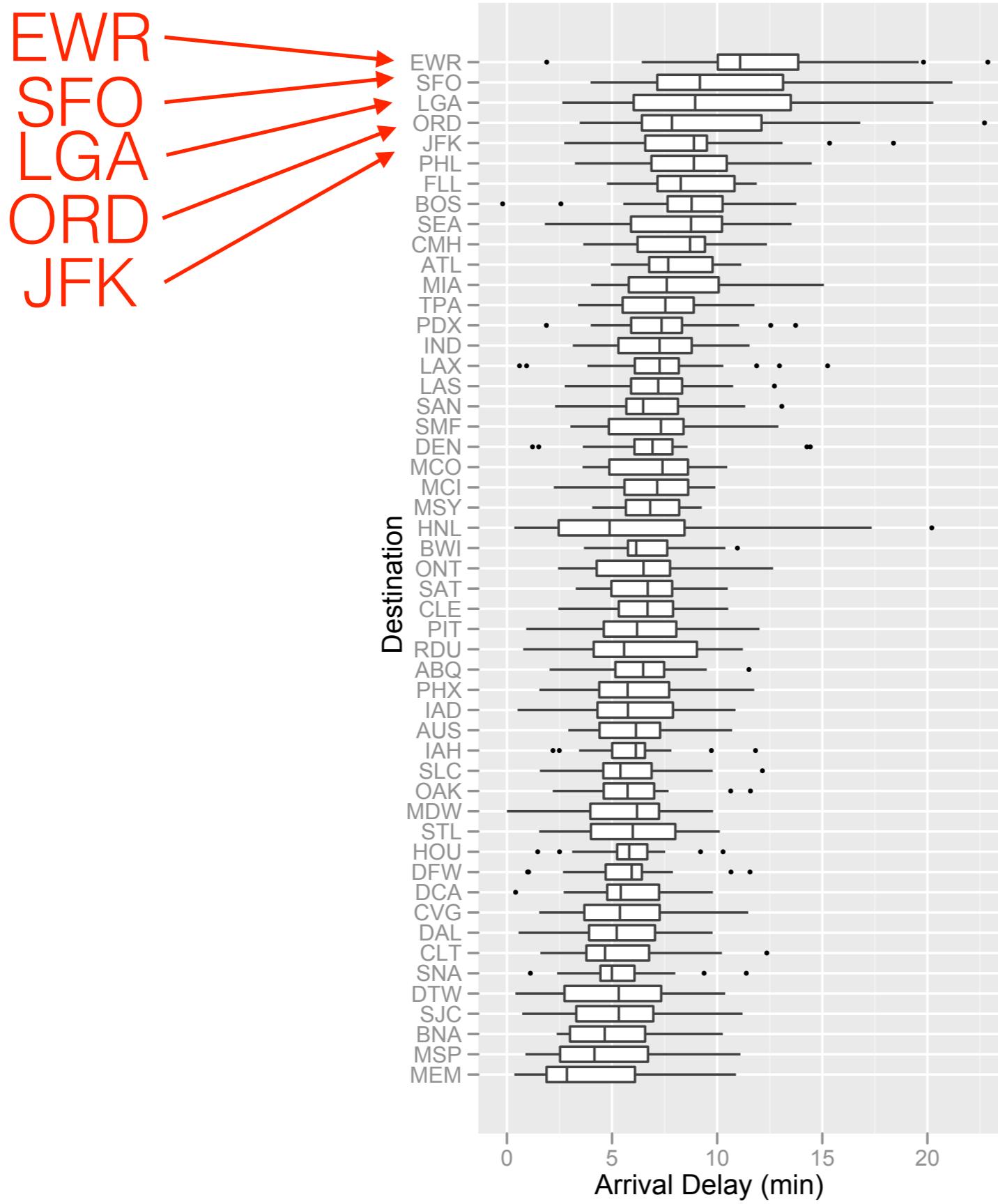


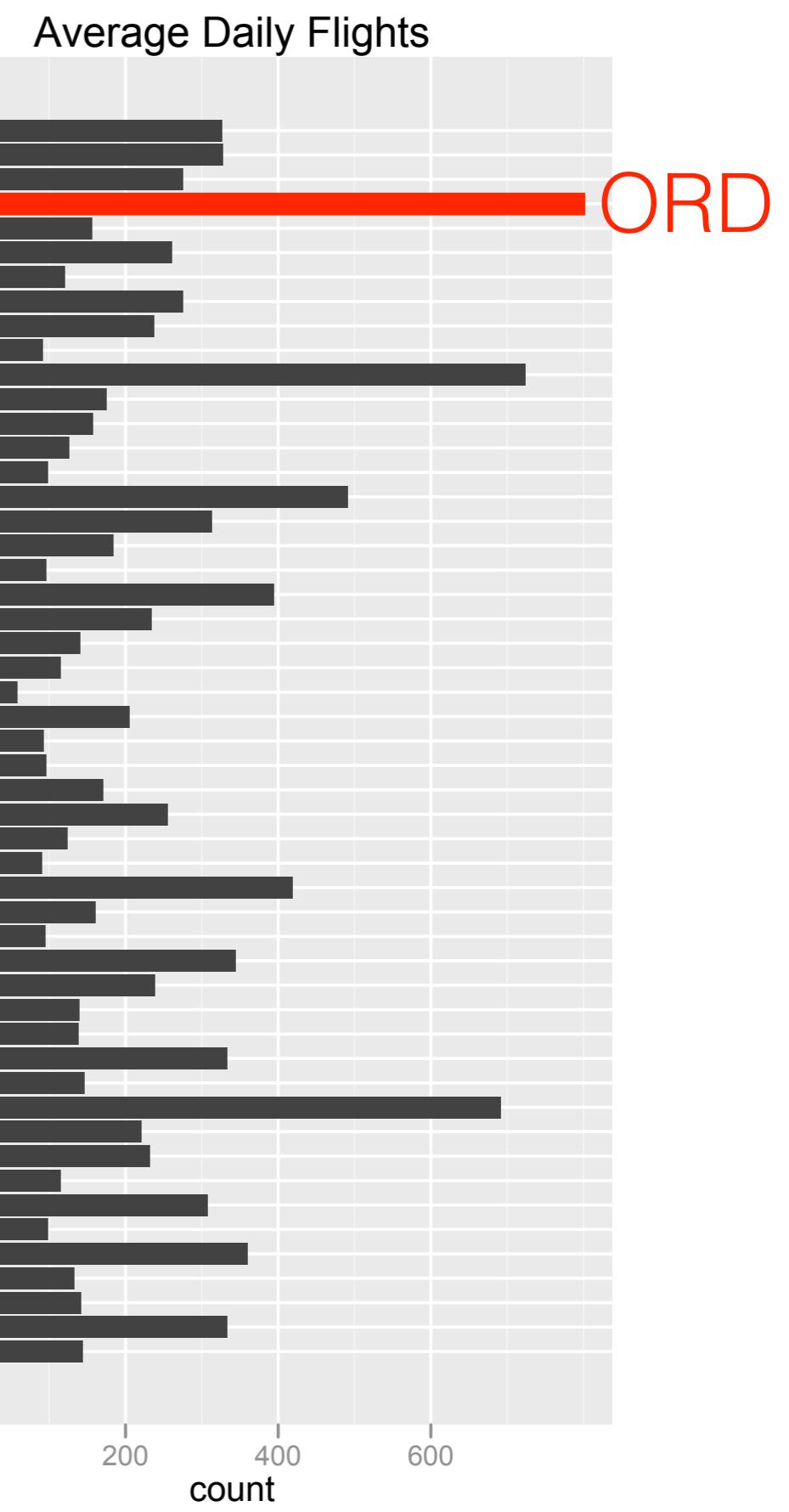
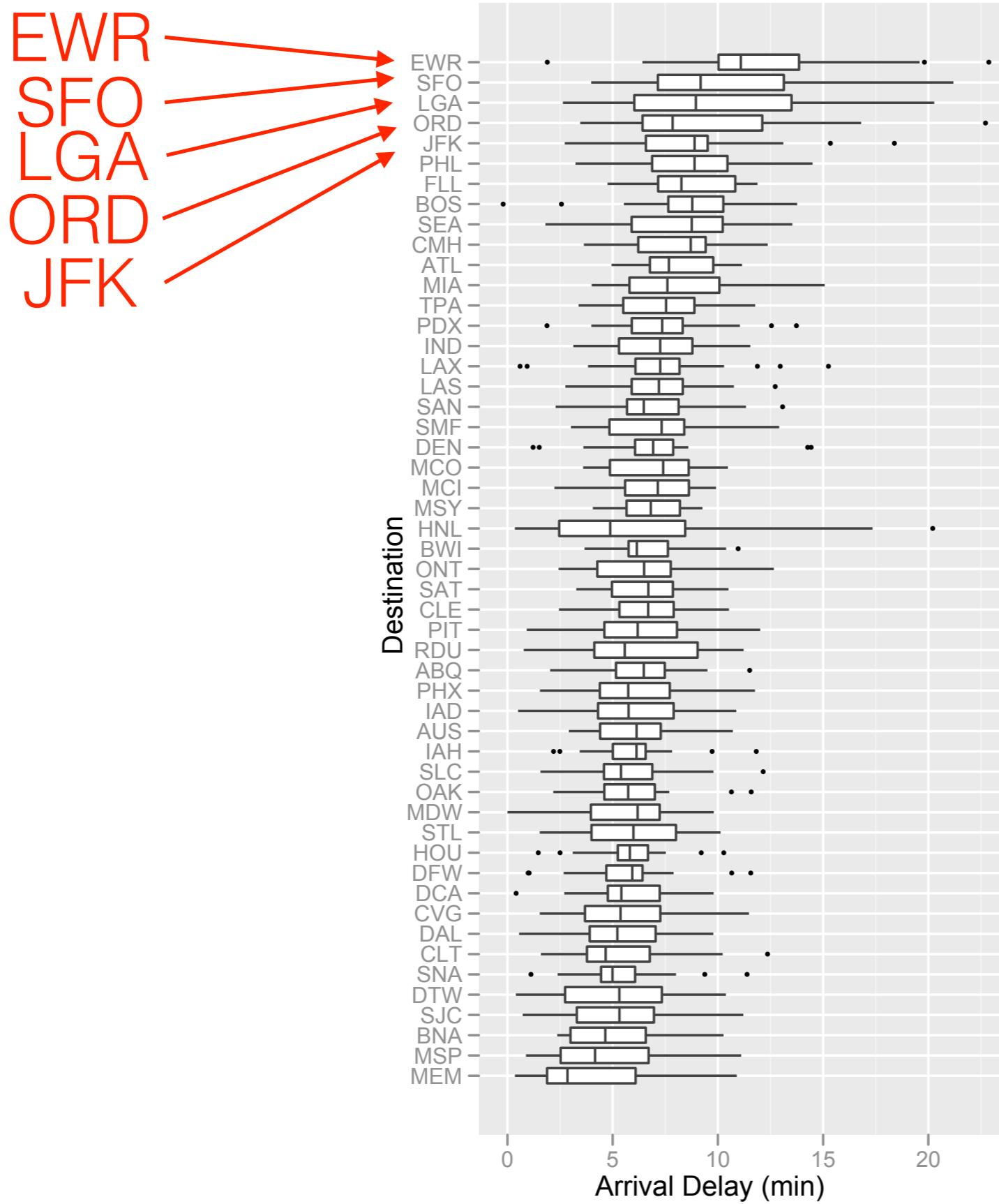




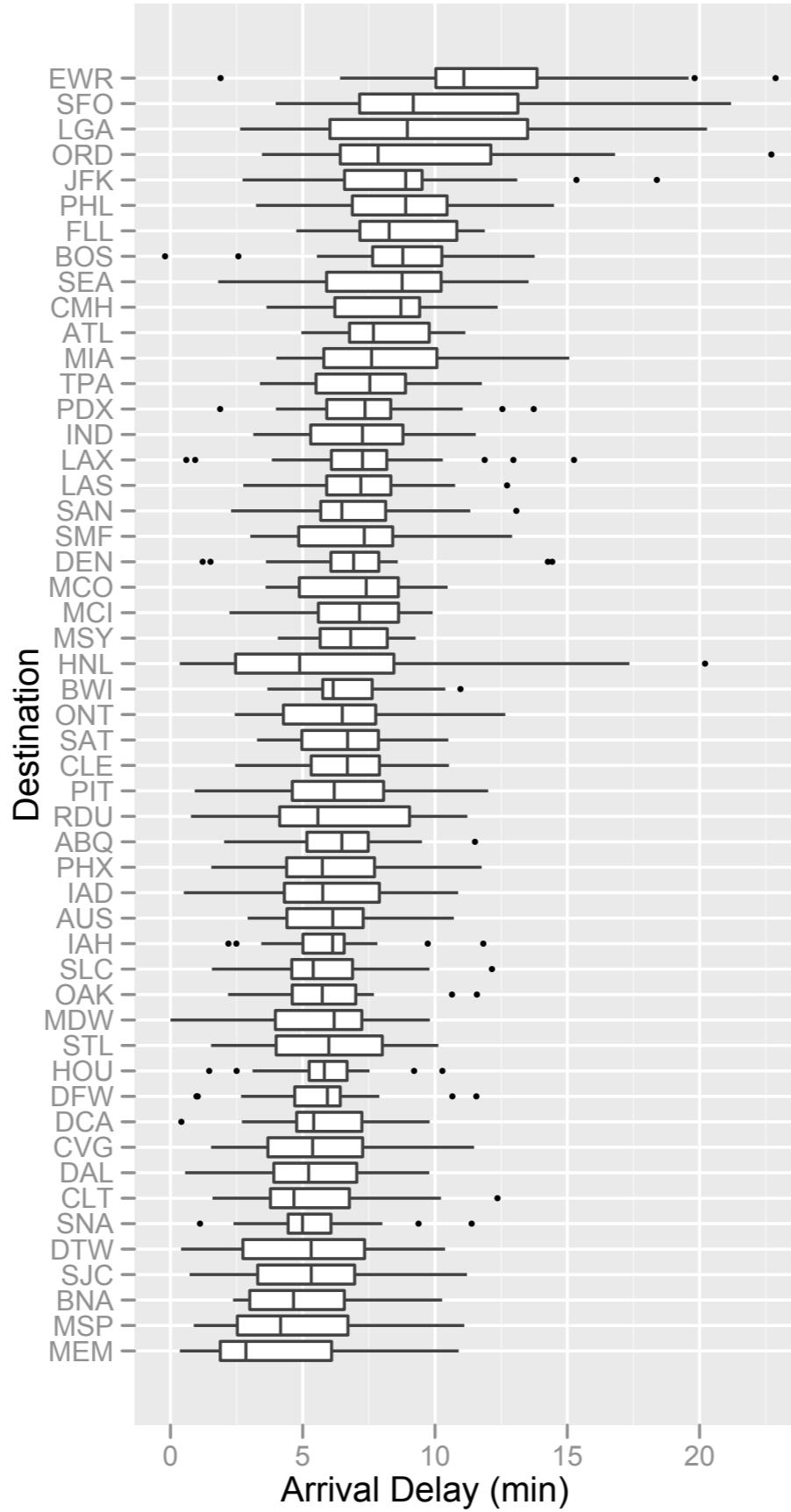
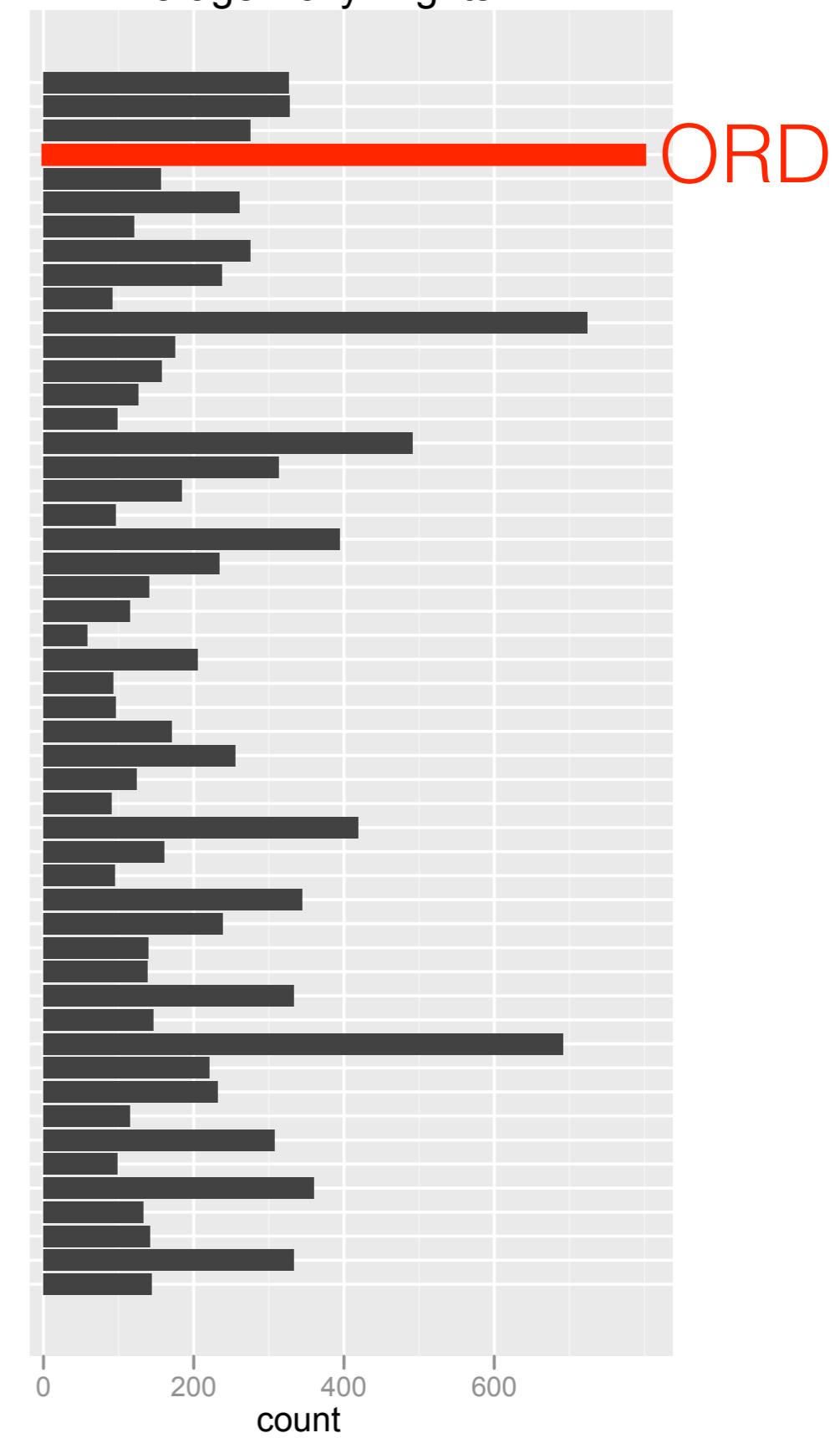
Average Daily Flights



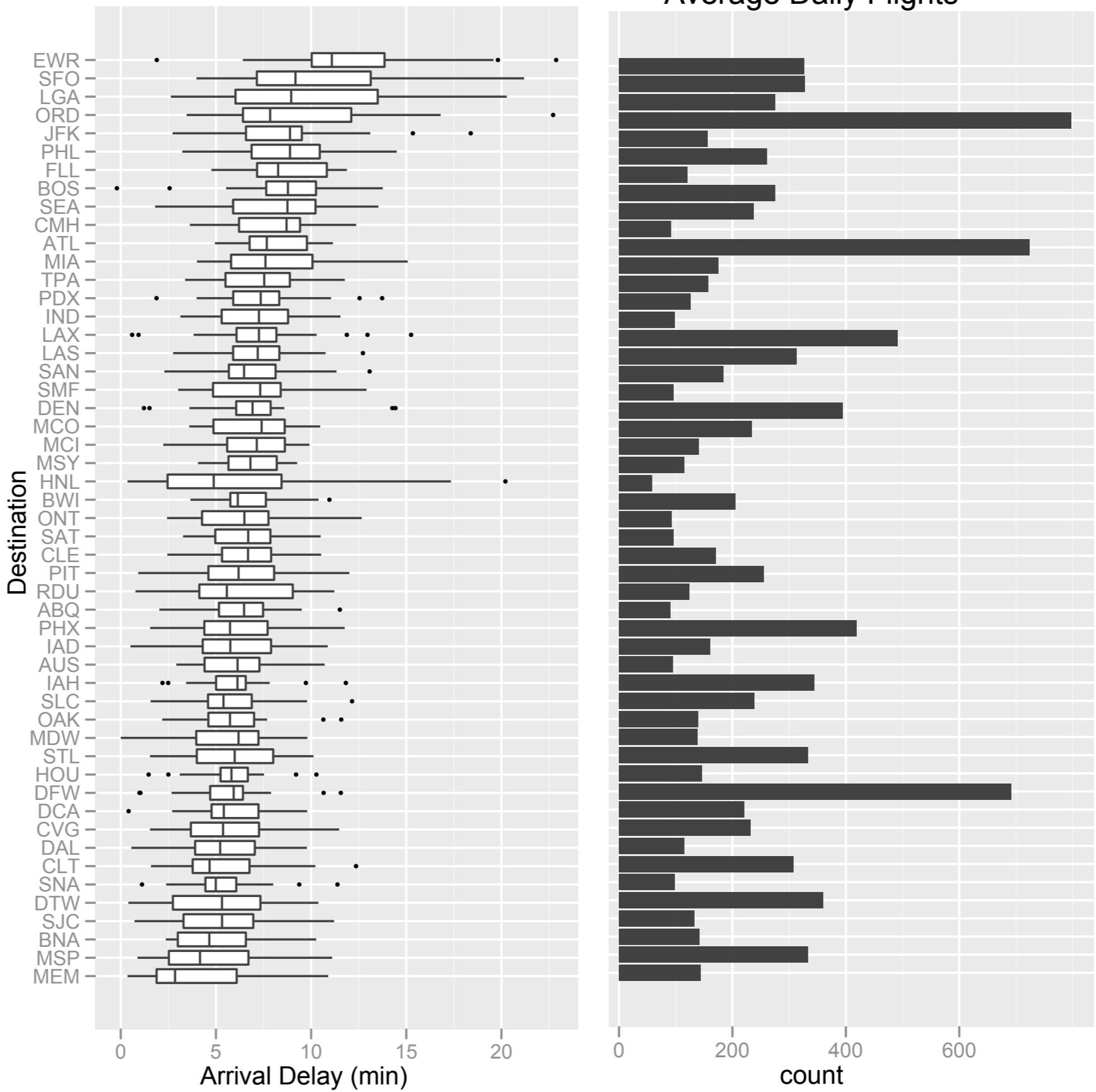




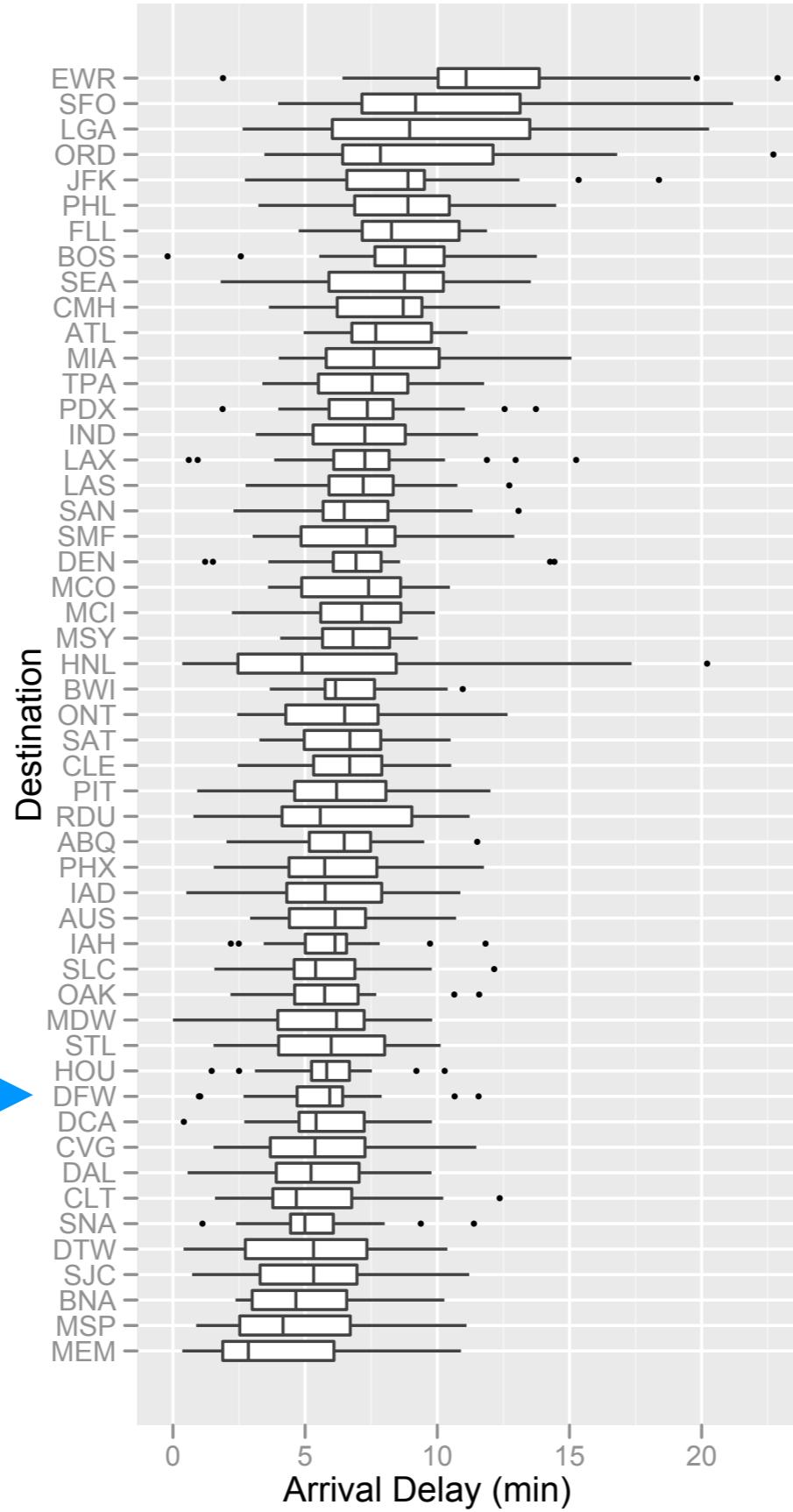
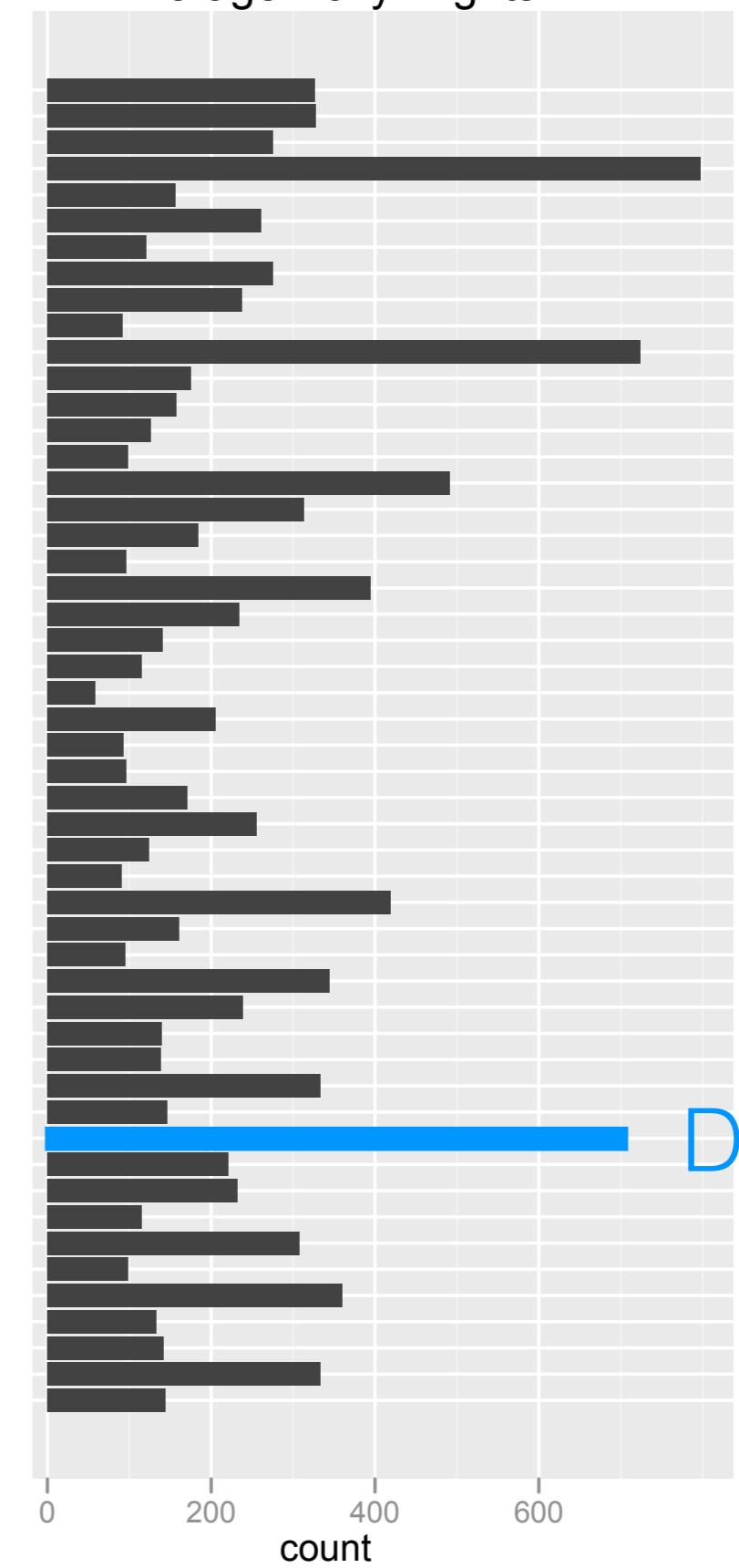
Average Daily Flights

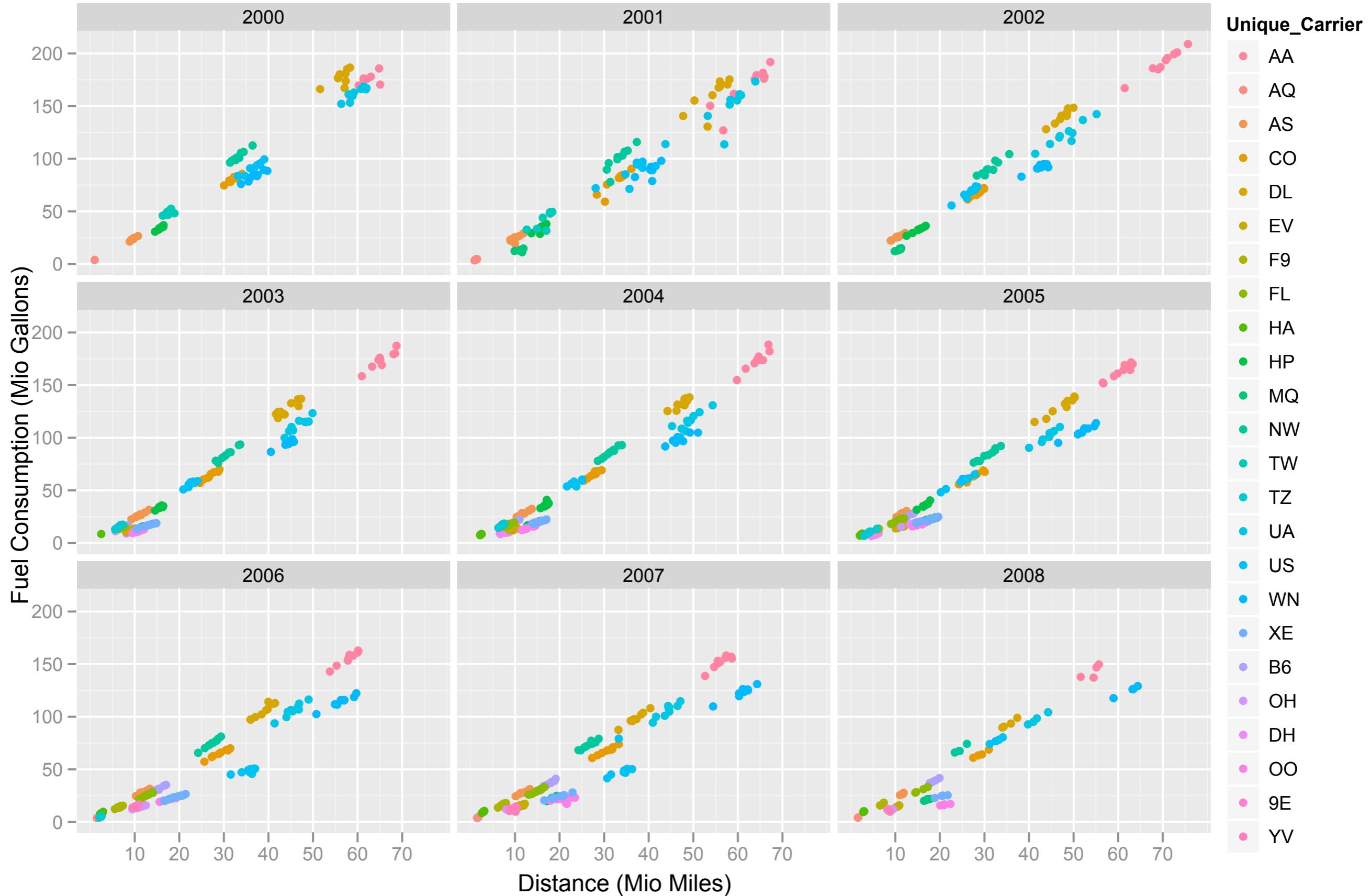


Average Daily Flights

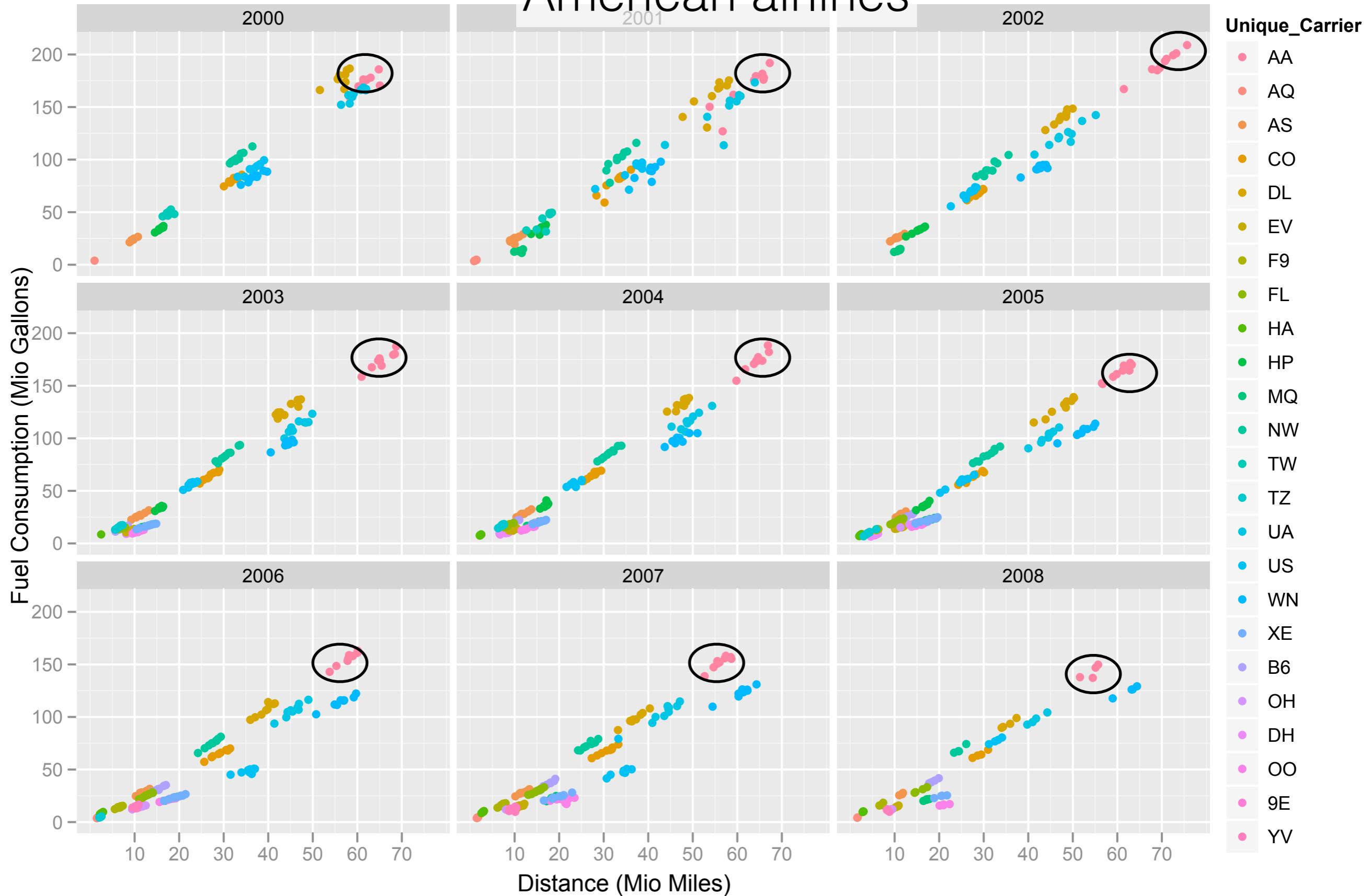


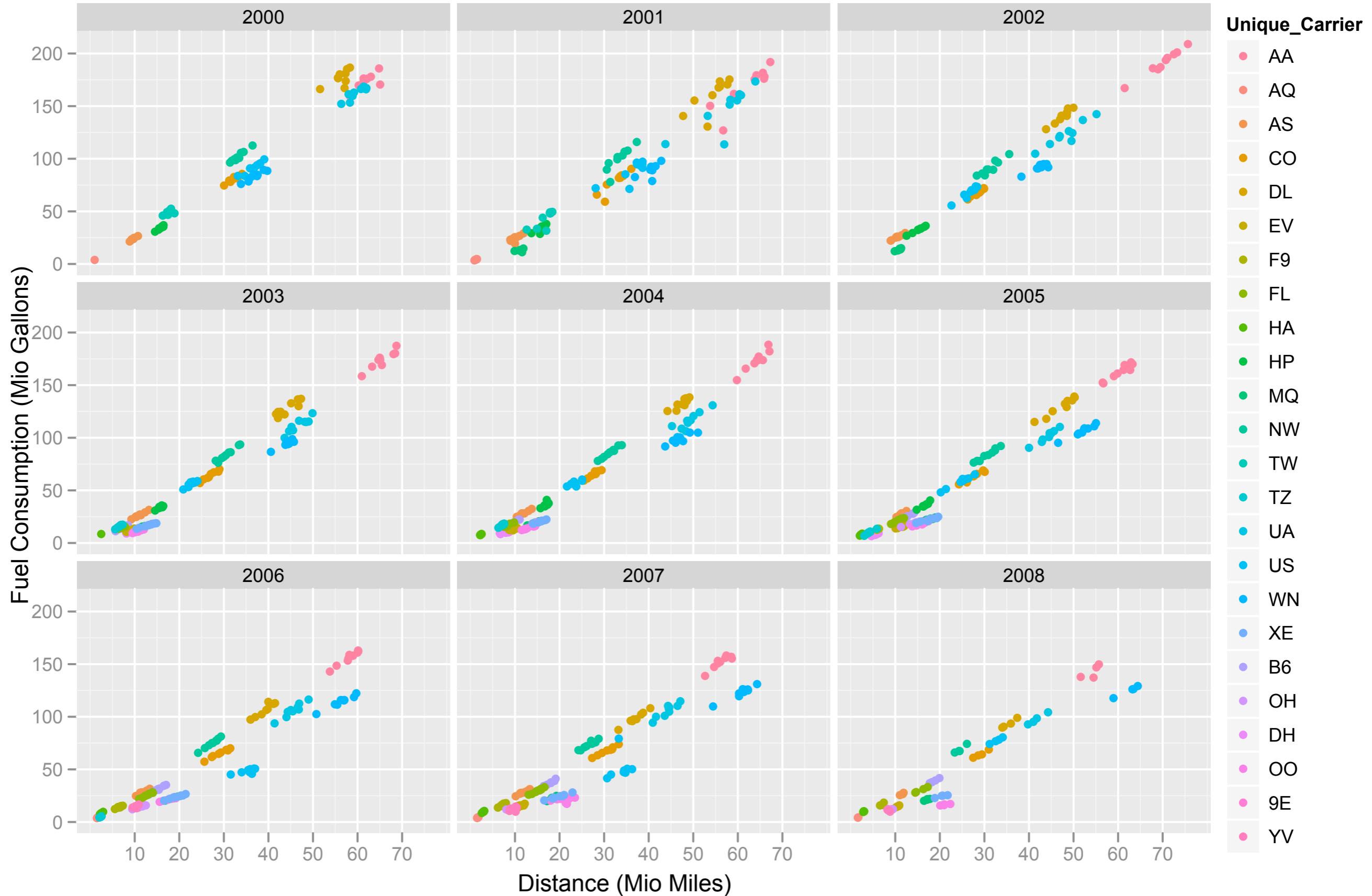
Average Daily Flights

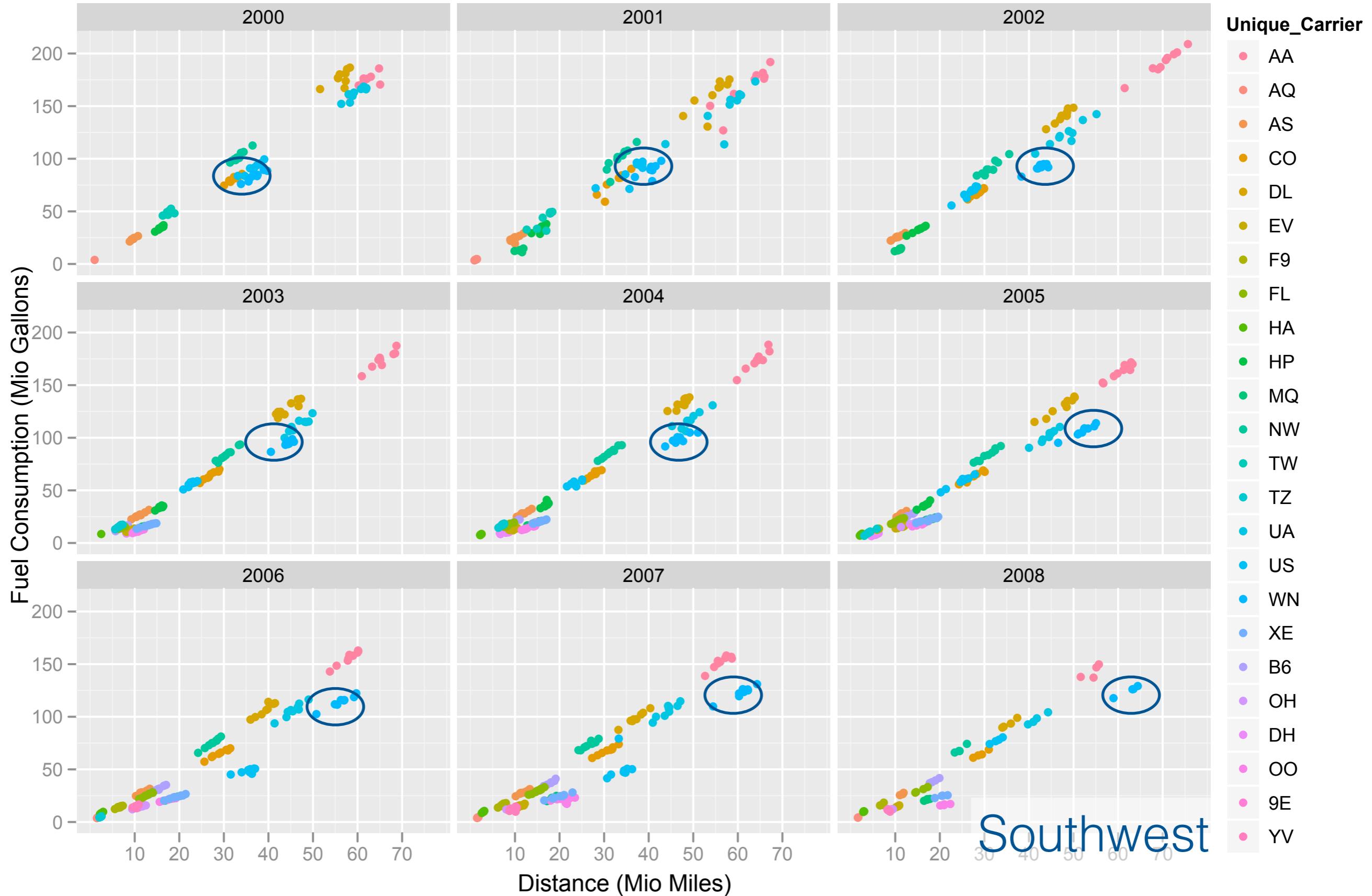


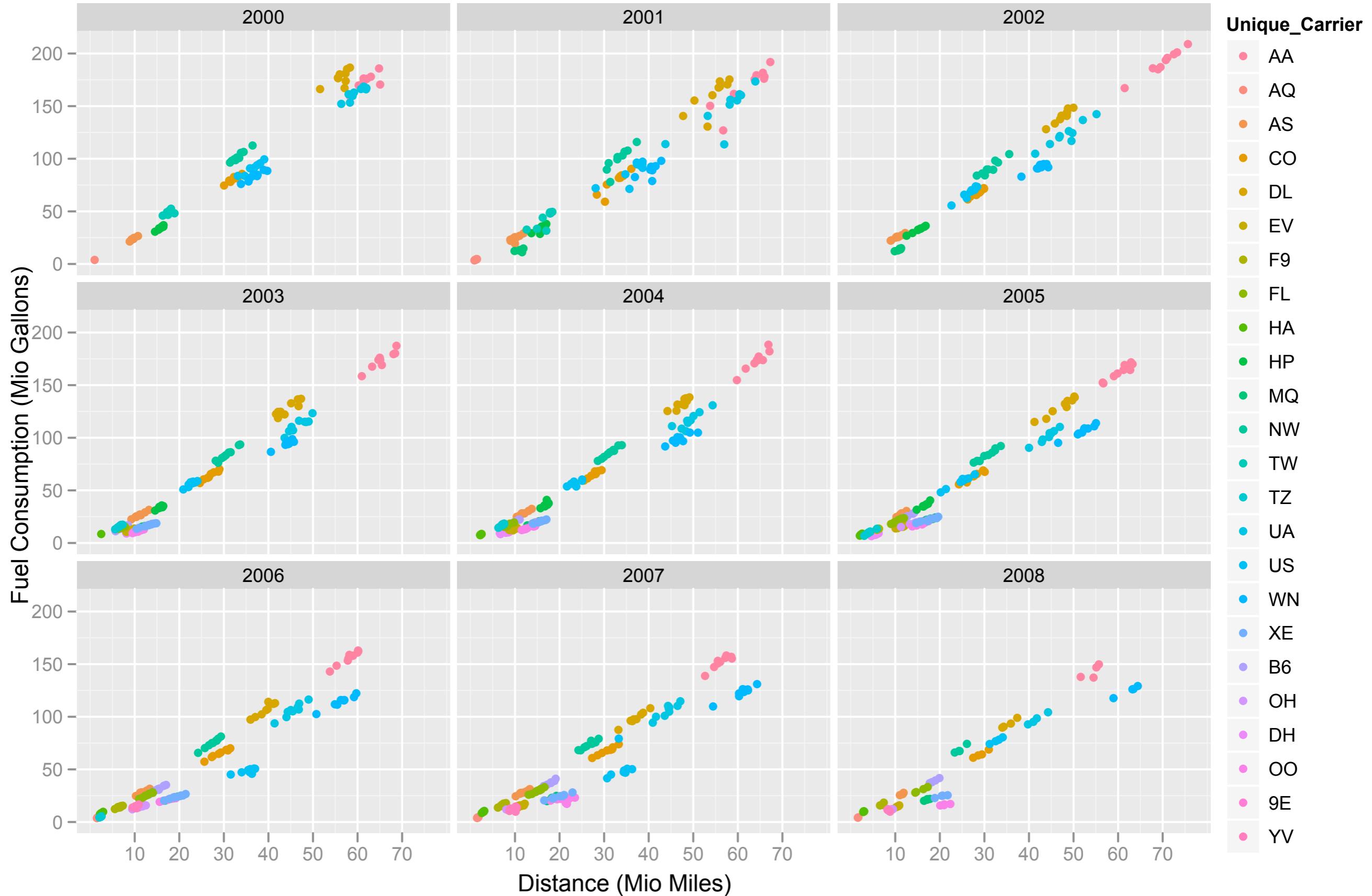


American airlines





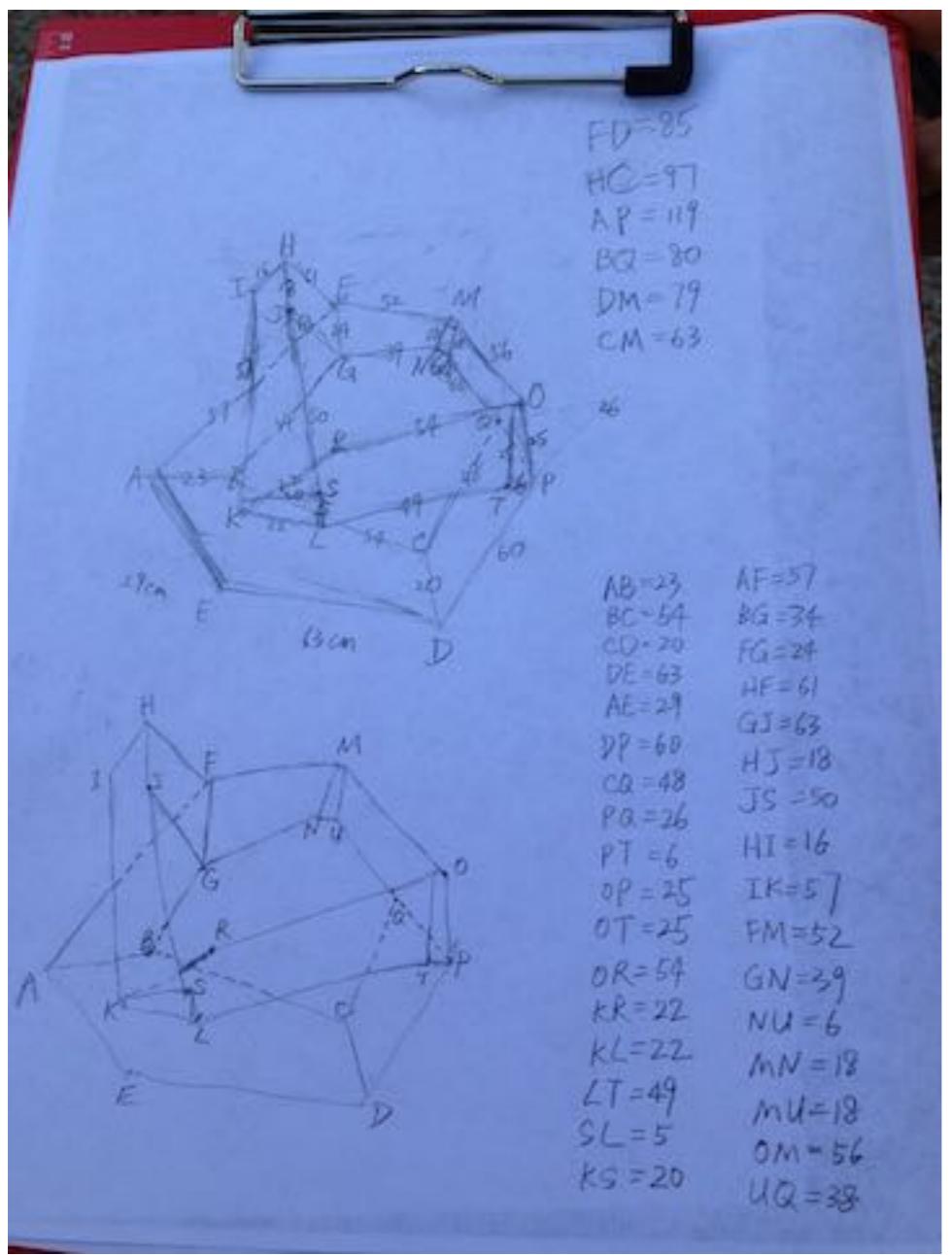




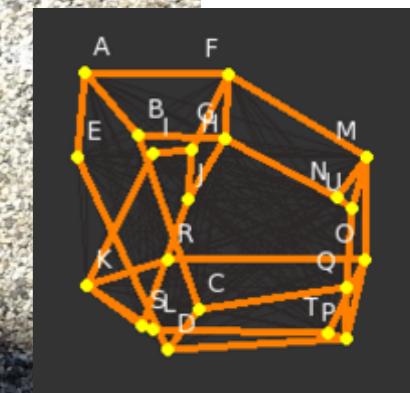
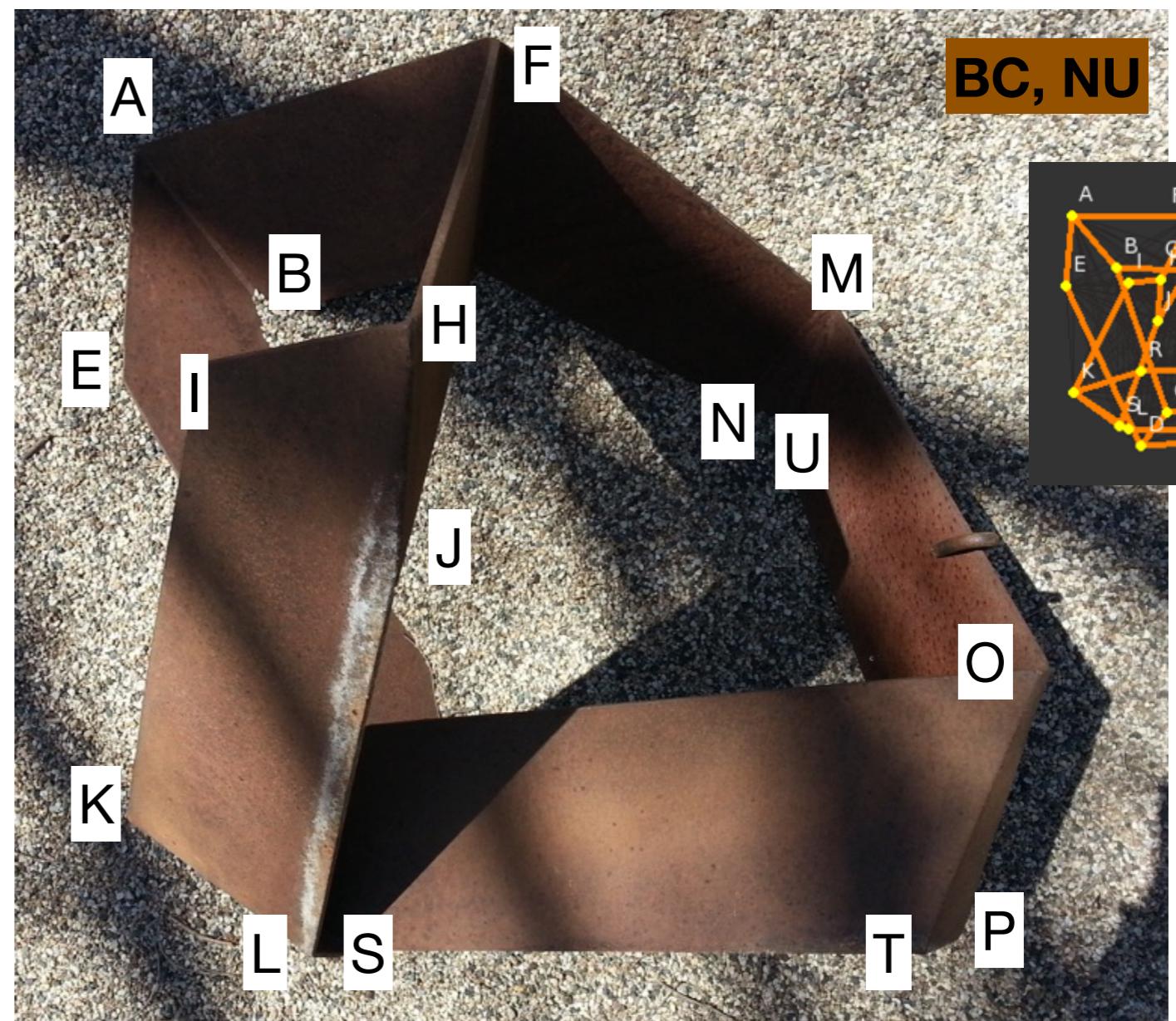
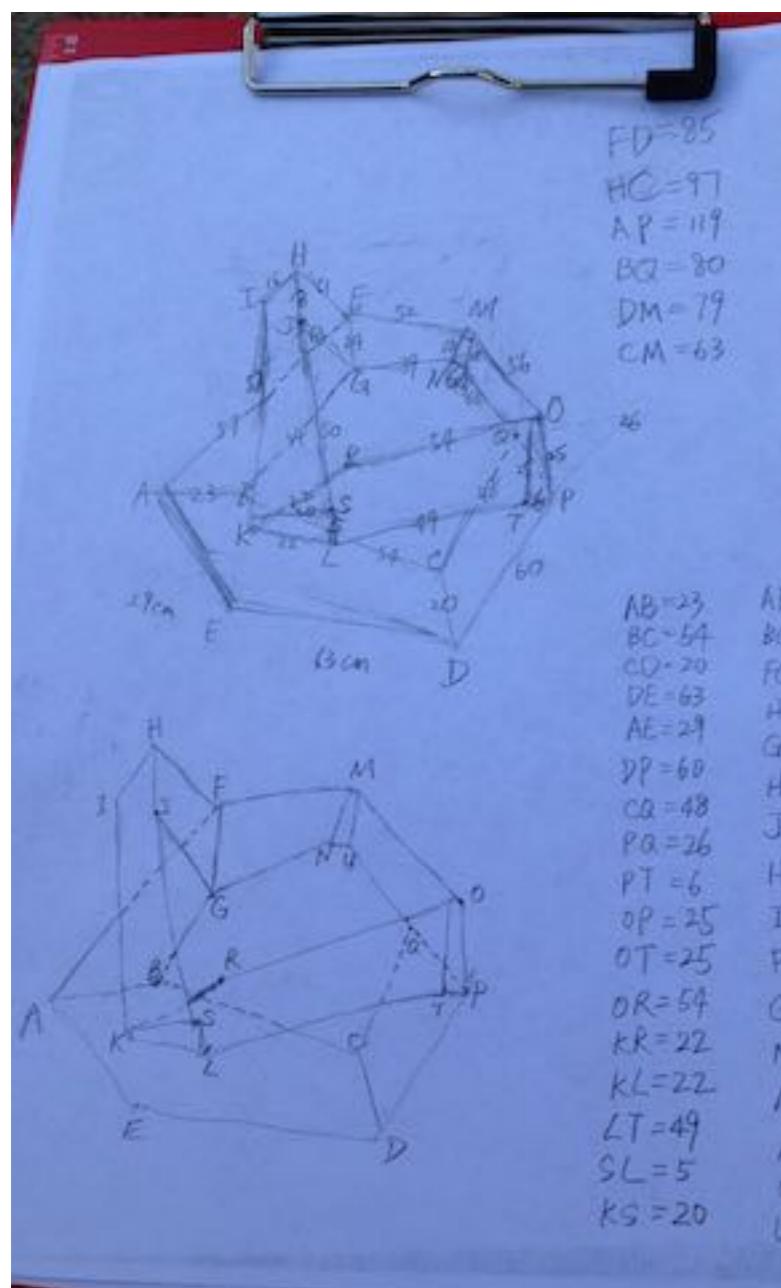
Sculpture exhibit

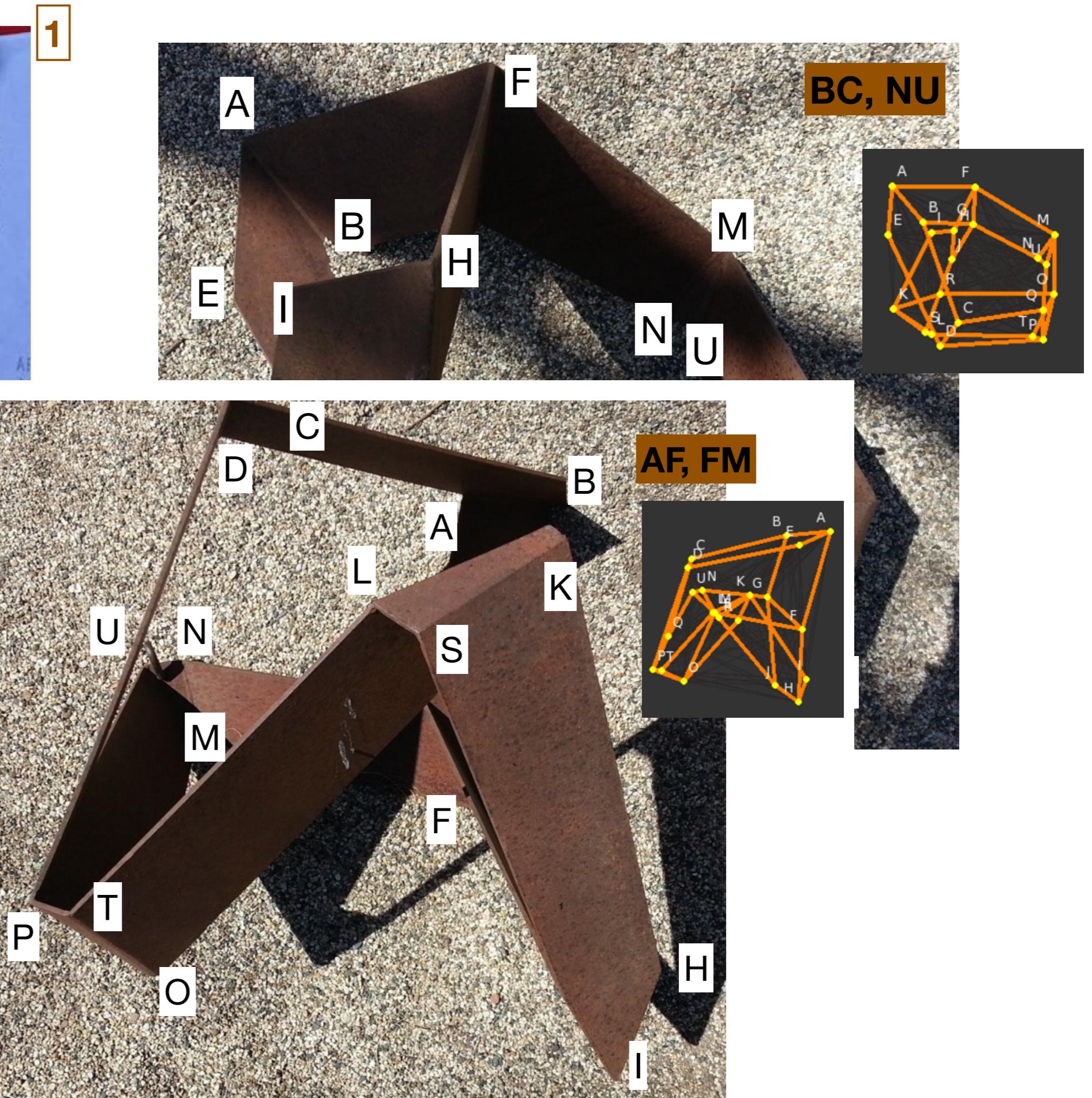
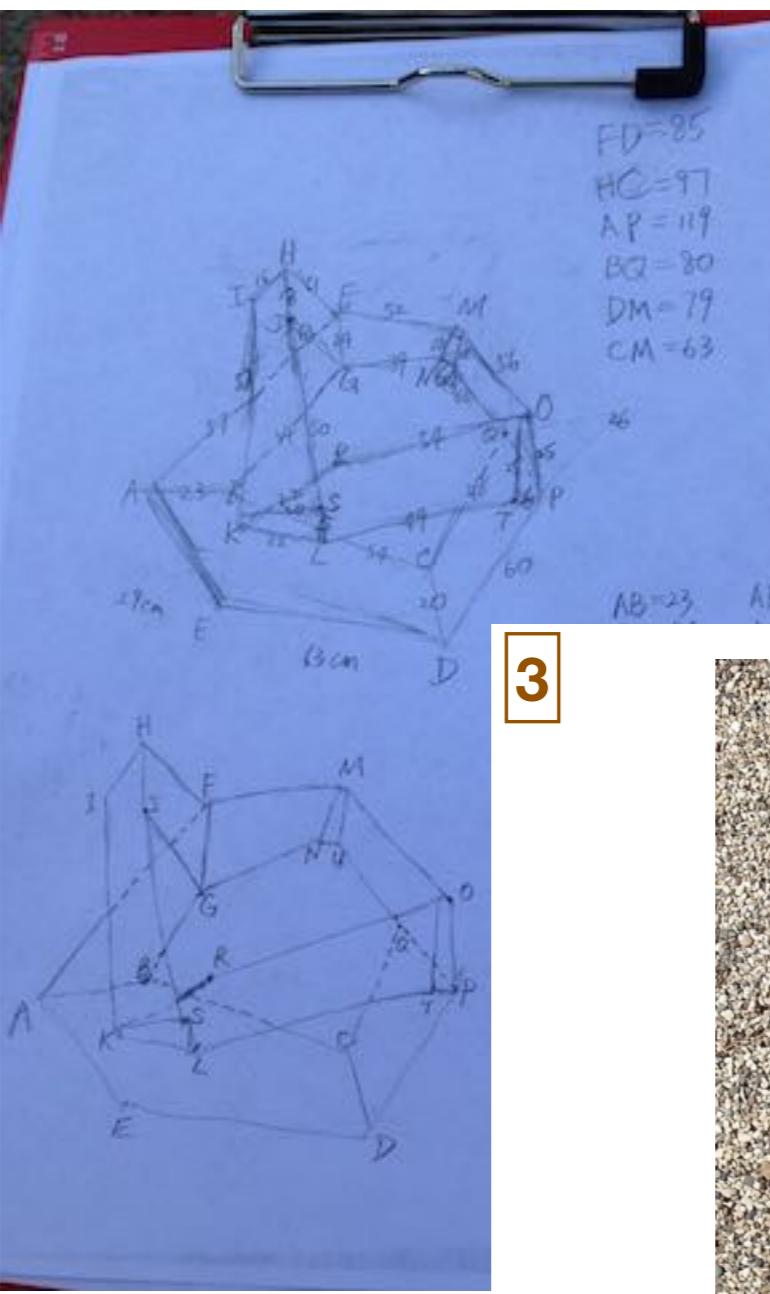


■ CHALLENGE:
Catalog 15 identical
sculptures of Chuck
Ginnever into 15
DIFFERENT
orientations



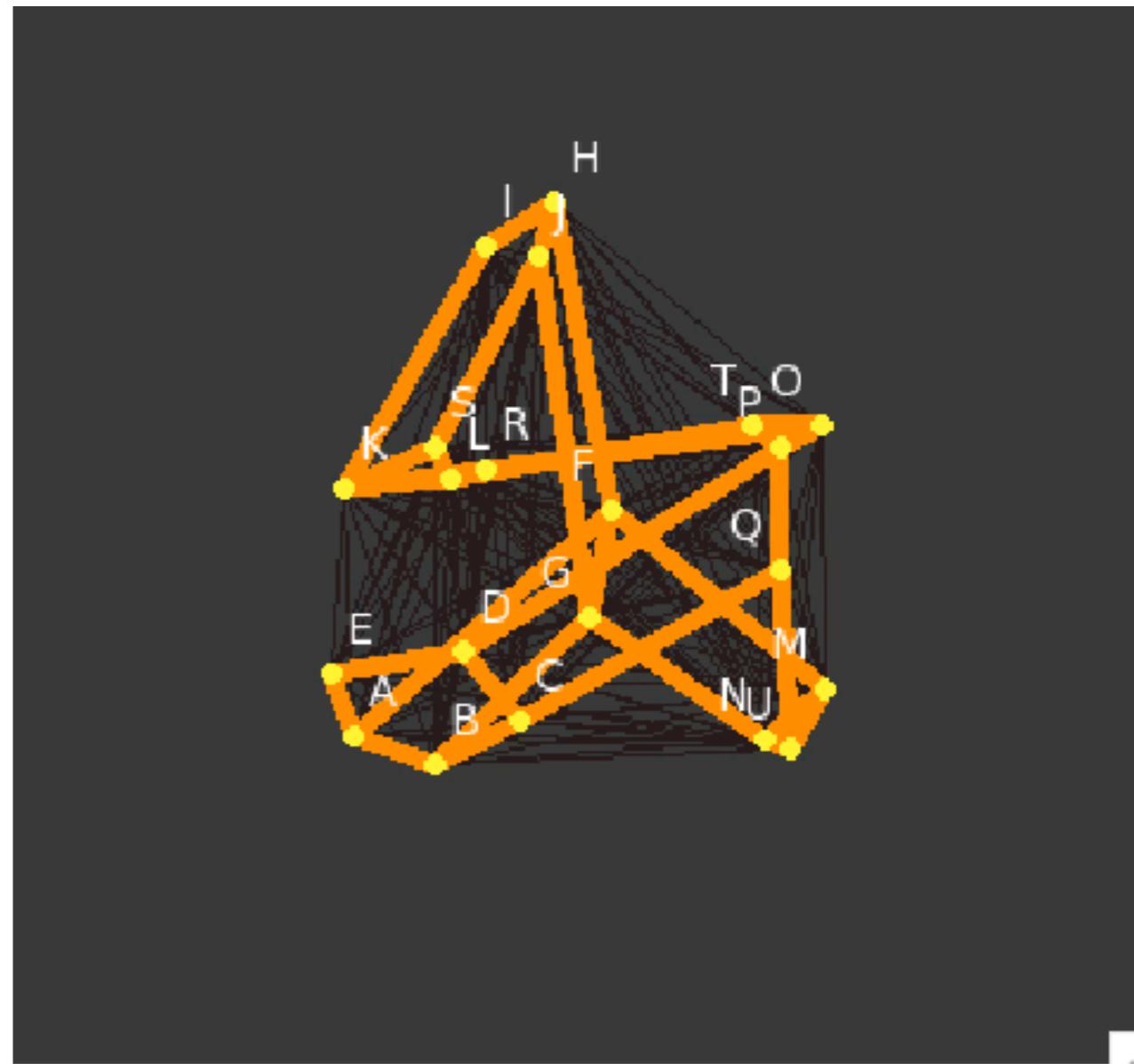
1





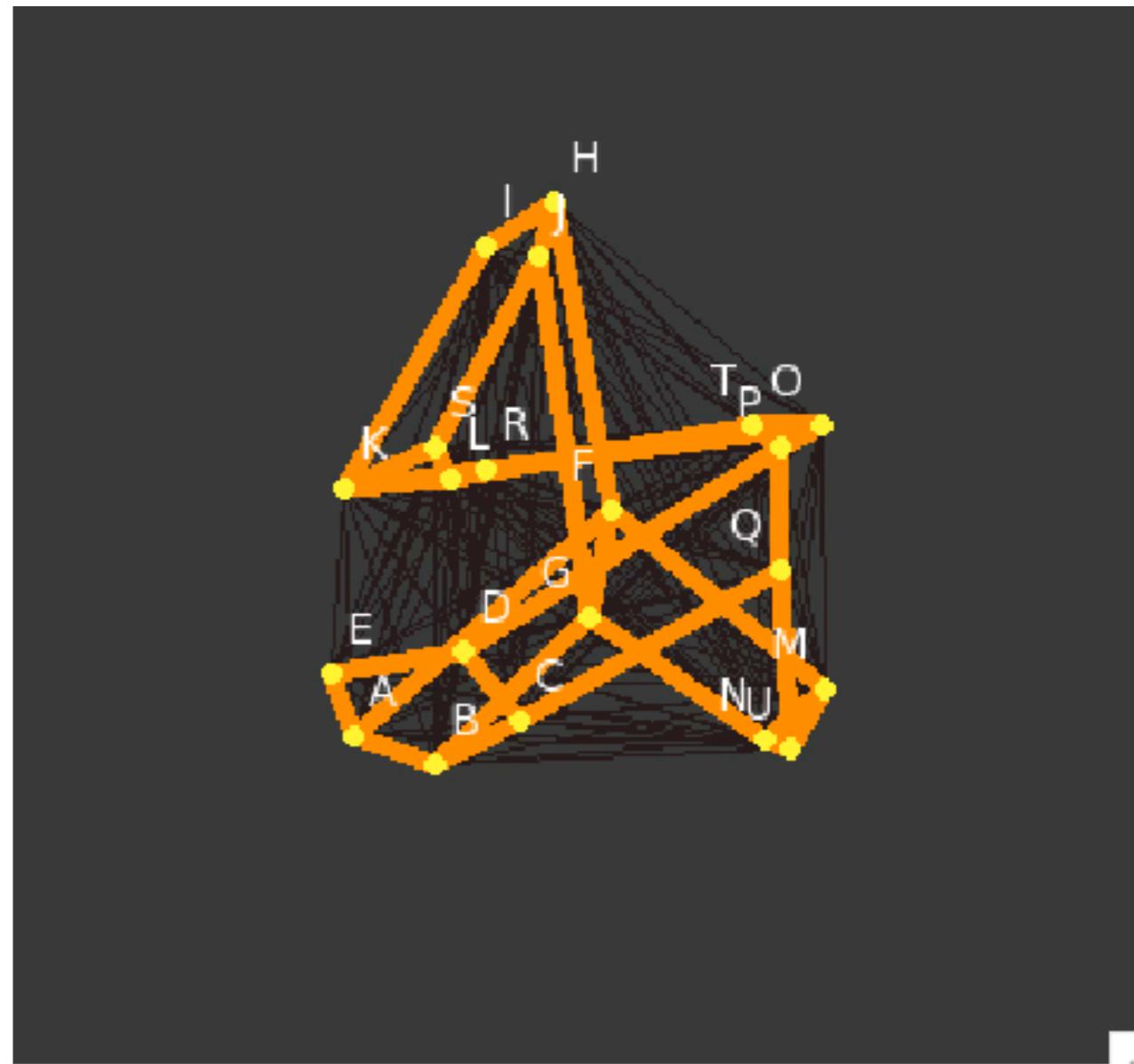
Sculpture exhibit

- Measured distances between vertices
- Used a method called multidimensional scaling to find a 3D representation that matched the distances
- Gave us a virtual model
- <https://vimeo.com/171196380>



Sculpture exhibit

- Measured distances between vertices
- Used a method called multidimensional scaling to find a 3D representation that matched the distances
- Gave us a virtual model
- <https://vimeo.com/171196380>



Today's project

- Becoming a business analyst requires math skills, and programming skills. There'll be a bit of both in the lab exercise.
- We will be working with the women's AFL statistics from the 2017 and 2018 seasons.



About the data

- For every player in 2017 and 2018, who actually played we have
 - ✓ Kicks
 - ✓ Handballs
 - ✓ Dispatch efficiency
 - ✓ Marks
 - ✓ Frees against
 - ✓ Goals
 - ✓ Goal assists
 - ✓ Time on ground
- Stats given are averages, per match



<http://www.afl.com.au/womens>

Question

- How do the players compare with each other?
- Are some players similar to each other, based on their statistics?

Simple approach

 Make a plot

 Examine relationship

 Check the players

 Only allows analysis one
or two statistics at a time

Simple approach

 Make a plot

 Examine relationship

 Check the players

 Only allows analysis one or two statistics at a time

Advanced approach

- Multidimensional scaling allows us to look at a combination of all the statistics, to examine how players are similar and different
- Compute the distance between all pairs of players using all the statistics

$$X = (x_1, x_2)$$

$$Y = (y_1, y_2)$$

$$d_{X,Y} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Advanced approach

- Multidimensional scaling allows us to look at a combination of all the statistics, to examine how players are similar and different
- Compute the distance between all pairs of players using all the statistics

Advanced approach

- Multidimensional scaling allows us to look at a combination of all the statistics, to examine how players are similar and different
- Compute the distance between all pairs of players using all the statistics

$$X = (x_1, x_2, \dots, x_p)$$

$$Y = (y_1, y_2, \dots, y_p)$$

$$d_{X,Y} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

With n players, you have nxn distances between players,

$$d_{i,j}, i = 1, \dots, n; j = 1, \dots, n; d_{i,i} = 0; d_{i,j} = d_{j,i}$$

We want to find a 2D arrangement where the distances between points best matches the distance between them in the p-dimensional space.

Minimise this

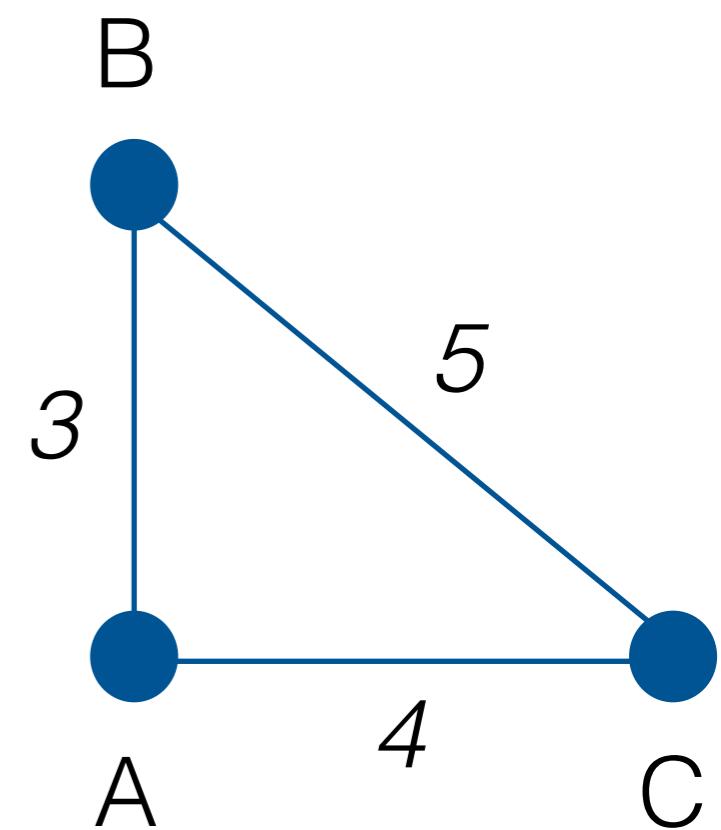
$$\text{Stress} = \sum_{i \neq j=1, \dots, n} (d_{ij} - d_{ij}^{2D})^2$$

	A	B	C
A	0	3	4
B	3	0	5
C	4	5	0

Draw a configuration of three points A, B, C, in 2D that matches this 3x3 distance matrix

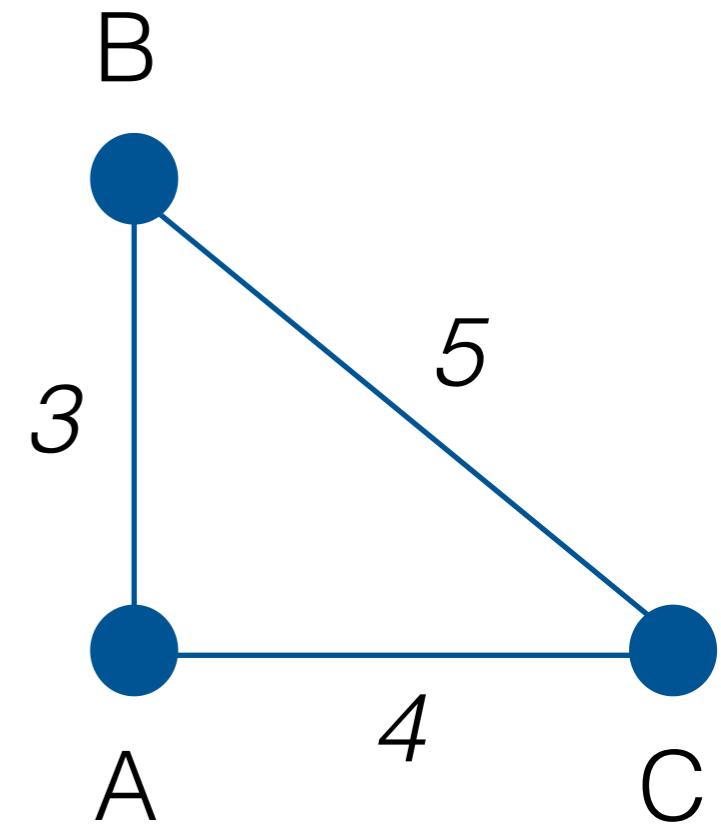
	A	B	C
A	0	3	4
B	3	0	5
C	4	5	0

Draw a configuration of three points A, B, C, in 2D that matches this 3x3 distance matrix



	A	B	C
A	0	3	4
B	3	0	5
C	4	5	0

Draw a configuration of three points A, B, C, in 2D that matches this 3x3 distance matrix

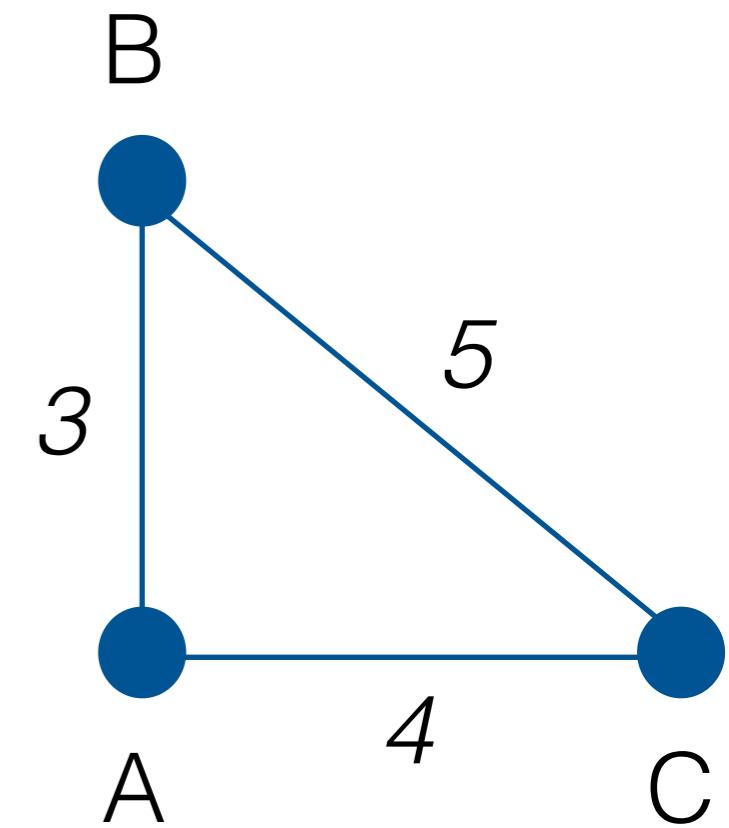


	A	B	C
A	0.0	3.1	4.2
B	3.1	0.0	5.1
C	4.2	5.1	0.0

Draw a configuration of three points A, B, C, in 2D that matches this 3x3 distance matrix

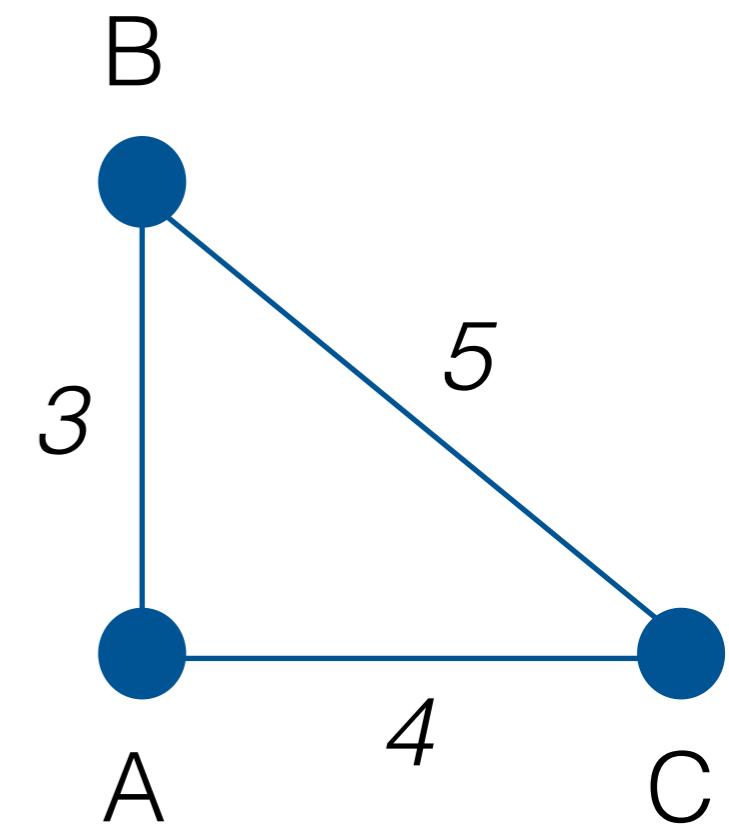
	A	B	C
A	0	3	4
B	3	0	5
C	4	5	0

Draw a configuration of three points A, B, C, in 2D that matches this 3x3 distance matrix

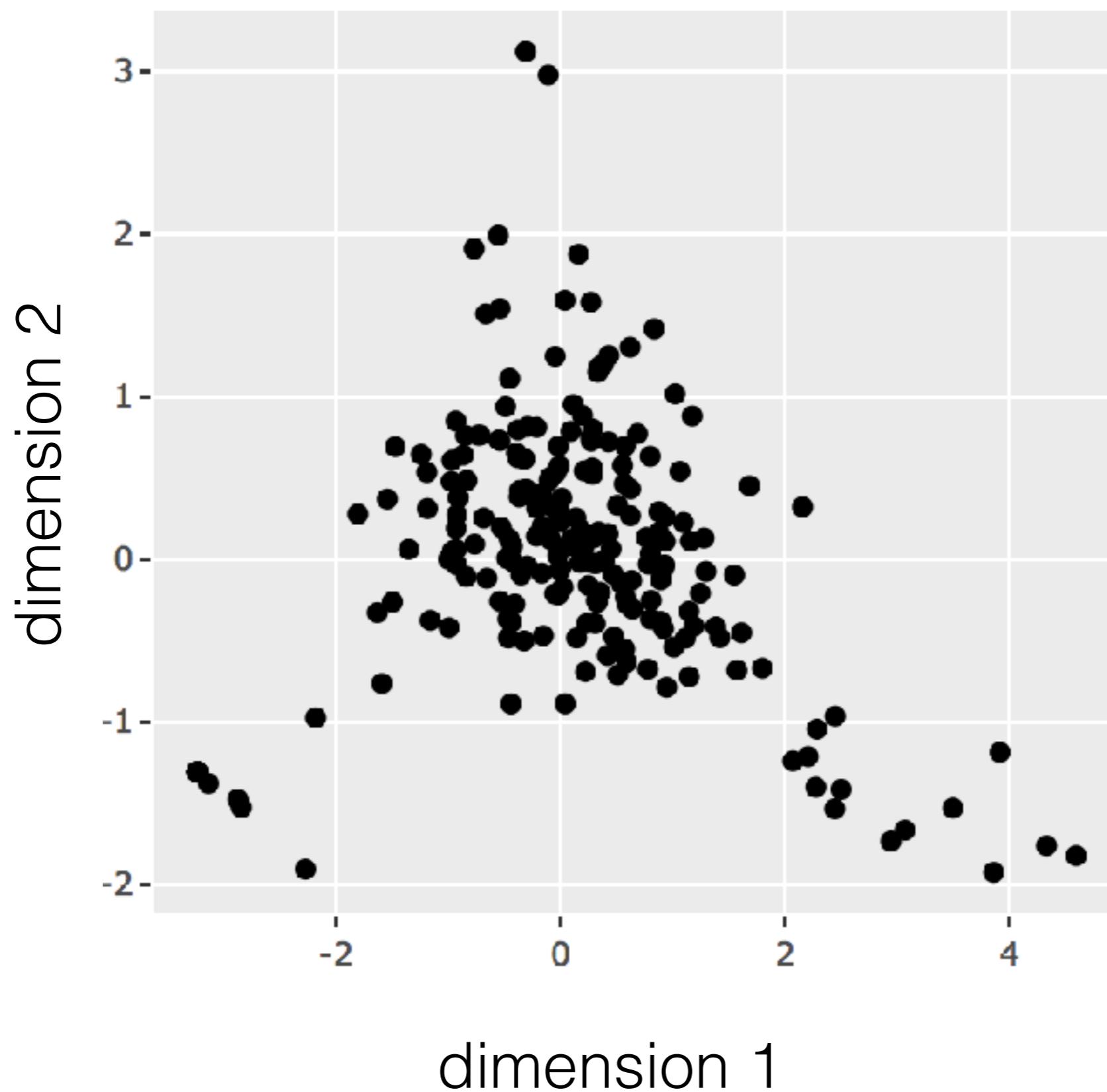


	A	B	C
A	0.0	3.1	4.2
B	3.1	0.0	5.1
C	4.2	5.1	0.0

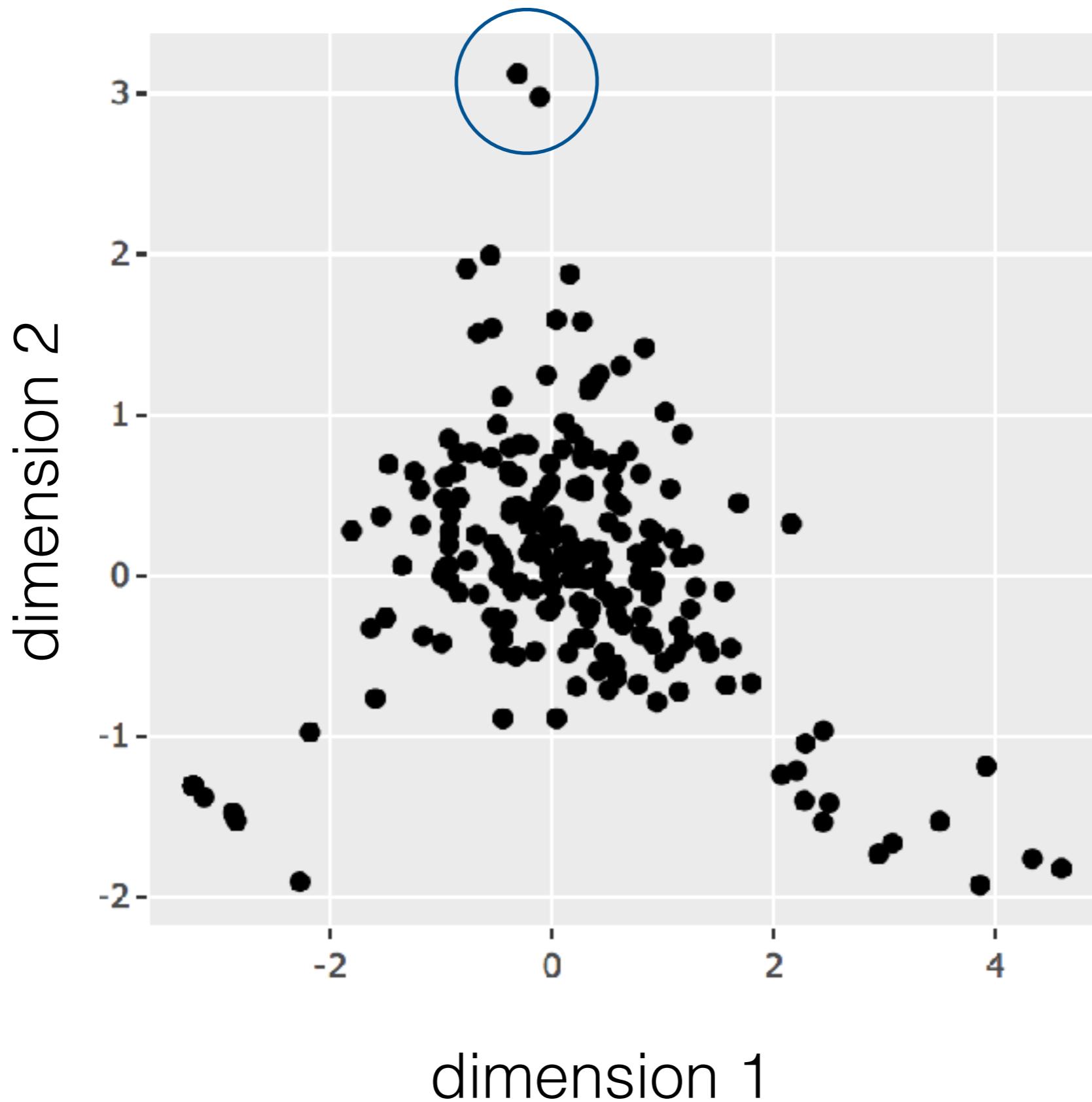
Draw a configuration of three points A, B, C, in 2D that matches this 3x3 distance matrix



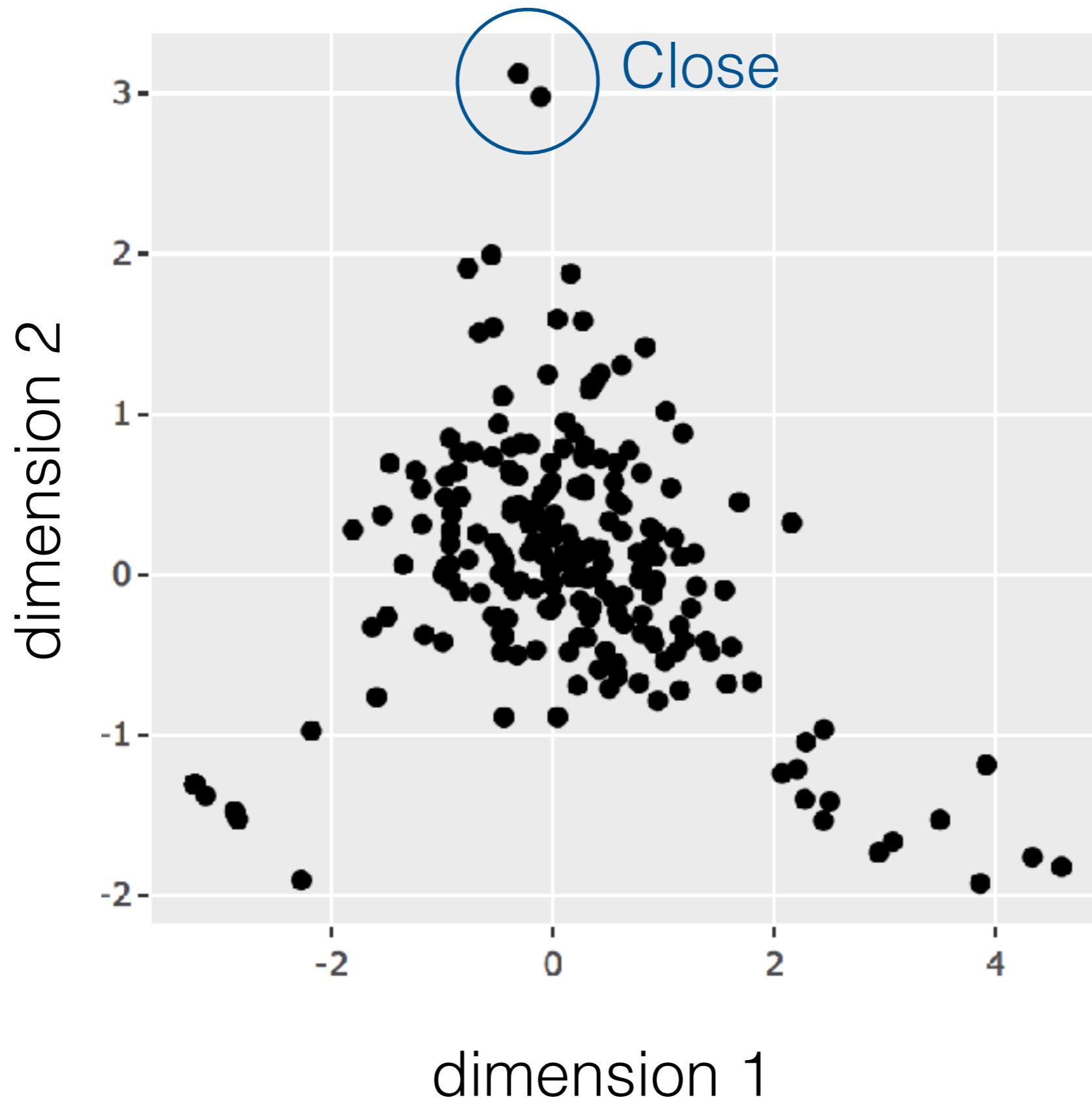
MDS plot



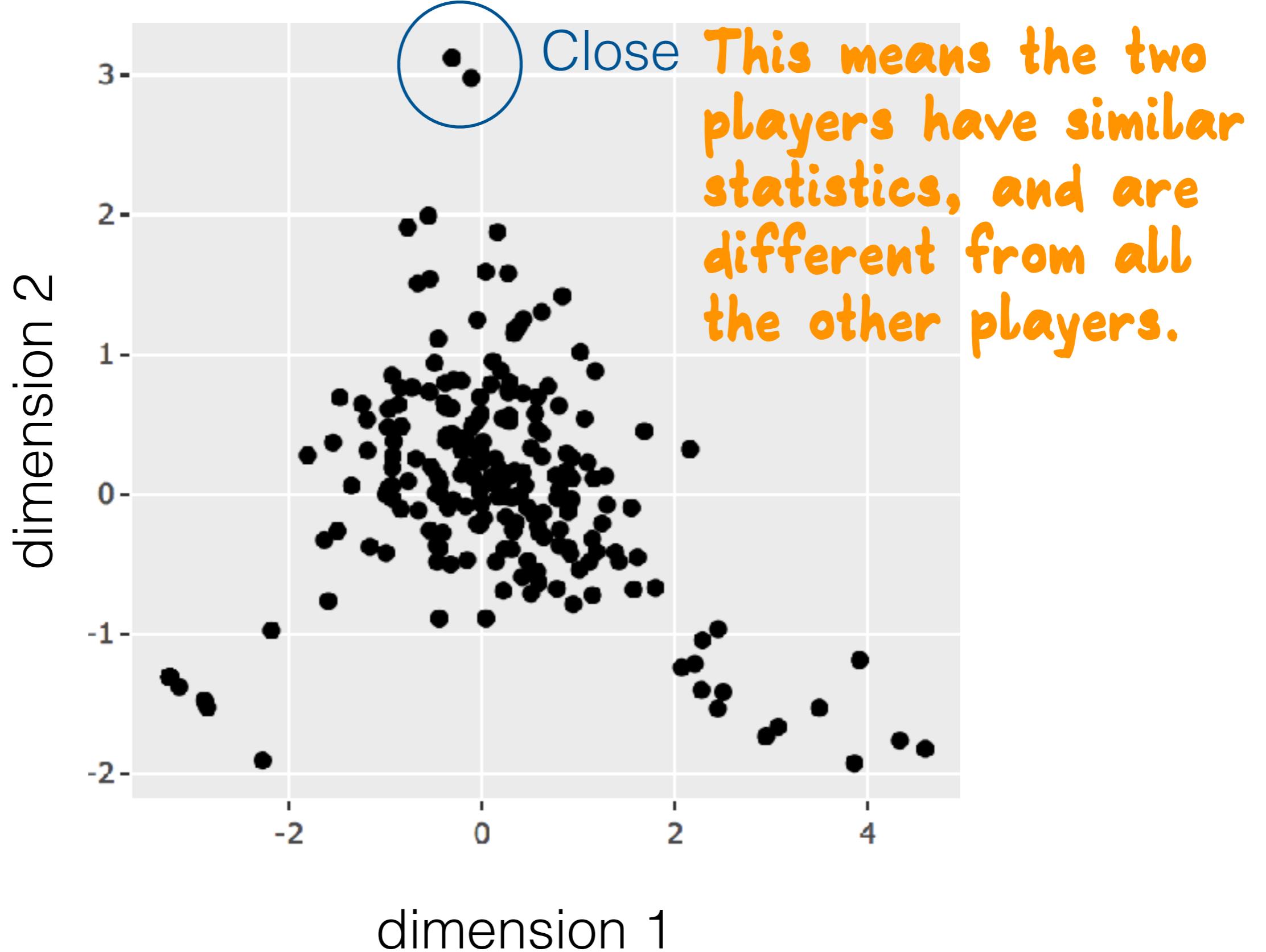
MDS plot



MDS plot



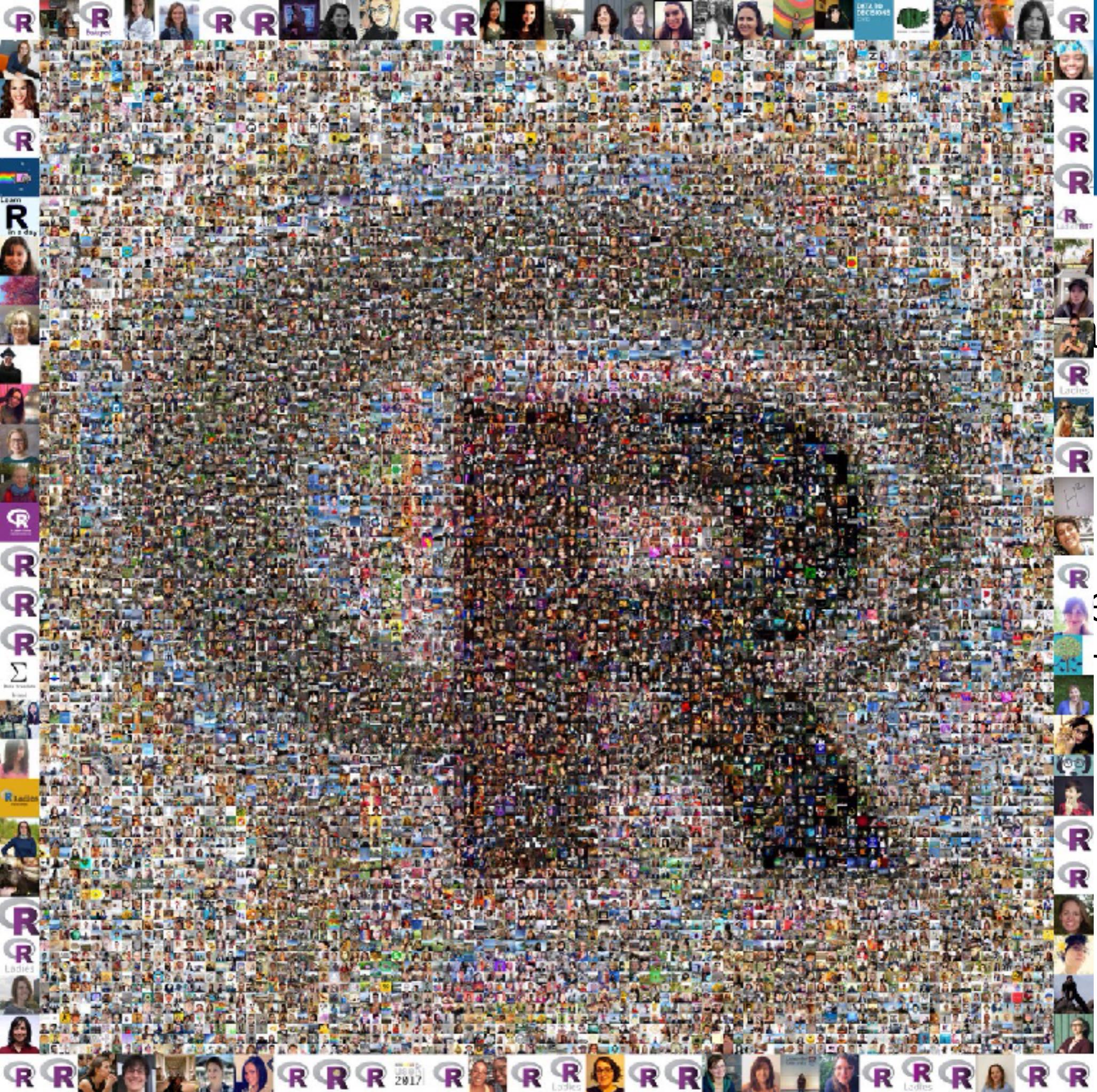
MDS plot



What is R?

- “R has become the most popular language for data science and an essential tool for Finance and analytics-driven companies such as Google, Facebook, and LinkedIn.” Microsoft 2015
- You name the company, pretty high probability they are using R for their big data analytics. Government departments, banks, local water resource management, are using R.
- R is taught in high school in New Zealand.

- “R” data
- sc
- an
- Fa
- You’re
are de-
made
- R ey
R nt





R is a scripting language

```
library(shiny)
library(tidyverse)
library(plotly)

players <- read_csv("data/players.csv")

ggplot(players,
       aes(x = Kicks, y = Goals,
            colour = Club)) +
  geom_point(alpha = 0.8) +
  facet_wrap(~Year, ncol=2)
```

R is a scripting language

```
library(shiny)  
library(tidyverse)  
library(plotly)
```

Load libraries

```
players <- read_csv("data/players.csv")
```

```
ggplot(players,  
       aes(x = Kicks, y = Goals,  
            colour = Club)) +  
  geom_point(alpha = 0.8) +  
  facet_wrap(~Year, ncol=2)
```

R is a scripting language

```
library(shiny)
library(tidyverse)
library(plotly)

players <- read_csv("data/players.csv")

ggplot(players,
       aes(x = Kicks, y = Goals,
            colour = Club)) +
  geom_point(alpha = 0.8) +
  facet_wrap(~Year, ncol=2)
```

R is a scripting language

```
library(shiny)
library(tidyverse)
library(plotly)
```

```
players <- read_csv("data/players.csv")
```

Read data

```
ggplot(players,
       aes(x = Kicks, y = Goals,
            colour = Club)) +
  geom_point(alpha = 0.8) +
  facet_wrap(~Year, ncol=2)
```

R is a scripting language

```
library(shiny)
library(tidyverse)
library(plotly)

players <- read_csv("data/players.csv")

ggplot(players,
       aes(x = Kicks, y = Goals,
            colour = Club)) +
  geom_point(alpha = 0.8) +
  facet_wrap(~Year, ncol=2)
```

R is a scripting language

```
library(shiny)
library(tidyverse)
library(plotly)

players <- read_csv("data/players.csv")
```

```
ggplot(players,
       aes(x = Kicks, y = Goals,
            colour = Club)) +
  geom_point(alpha = 0.8) +
  facet_wrap(~Year, ncol=2)
```

Plot data

R is a scripting language

```
library(shiny)
library(tidyverse)
library(plotly)

players <- read_csv("data/players.csv")

ggplot(players,
       aes(x = Kicks, y = Goals,
            colour = Club)) +
  geom_point(alpha = 0.8) +
  facet_wrap(~Year, ncol=2)
```

What is a package?

- The strength of R is that people across the globe contribute their own work in the form of packages
- Install a package once using:
`install.packages("package_name")`
- Use the package in any future session by loading it with:
`library("package_name")`

Pipes and “gets”

- “`<-`” is called “gets” and is actually the same as “`=`”
- “`%>%`” is called a pipe and belongs to the “magrittr” package. It allows various functions or commands to be sequentially run like a pipeline.

What is a function?

$$f(x) = \frac{1}{n} \sum_{i=1}^n w_i x_i$$

```
wgt_mean <- function(x, w) {  
  xbar <- sum(x*w)/nrow(x)  
  return(xbar)  
}
```

Prompts

In the R console window

- “>” means that it is waiting for you to tell it what to do next
- “+” means that you have started a command, but its not completely finished

“()”, “[]” and “{ }”

- () are for functions and commands

```
read_csv("players.csv")
```

- [] are for indexing data frames, vectors and matrices

```
m_e[1:2, ]
```

- { } delineate the body of a function

```
wgt_mean <- function(x, w) {  
  ...  
}
```

Today's lab

- Some background googling
- Read the data, open the web app, do some analysis of player statistics
- Check if the claims in two news articles is verified by the data
- Add a new tab to the web app to explore teams
- Make a slide with your findings

Power to the People

H. G. Wells (1903) Mankind in the Making

“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write!”

A combination of open data, open source software, and quantitative skills gives us the power to make a difference in this world.

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.