Apoyando el diseño y desarrollo de nuevas políticas públicas en educación usando Data Science

Dr. Patricio Rodríguez

prodriguez@ciae.uchile.cl

RedClara

Webminar "El potencial de la analítica avanzada en las Universidades"

Diciembre 02, 2015

Centro de Investigación Avanzada en Educación
Universidad de Chile

UNIVERSIDAD ADOLFO IBÁÑEZ

CONICYT Ministerio de Educación
Gobierno de Chile

FONDEF
Fondo de Fomento al Desarrollo Científico y Tecnológico

Red CLARA
+ Red + Ciencia

1 Desafíos

© 2015 — Todos los derechos reservados

1

Desafíos

Reforma educacional

Chile está en medio de una reforma educativa cuyos principales objetivos son:

1. Priorizar y fortalecer la educación pública, transfiriendo la administración desde los municipios a **Servicios Locales de Educación** (SLE) **Pública**
2. Regular la provisión de la oferta de los establecimientos particulares subvencionados: eliminación de selección, lucro y co-pago
3. Fortalecer la profesión docente

© 2015 — Todos los derechos reservados

1 Desafíos

Diversidad

Chile es un país diverso:

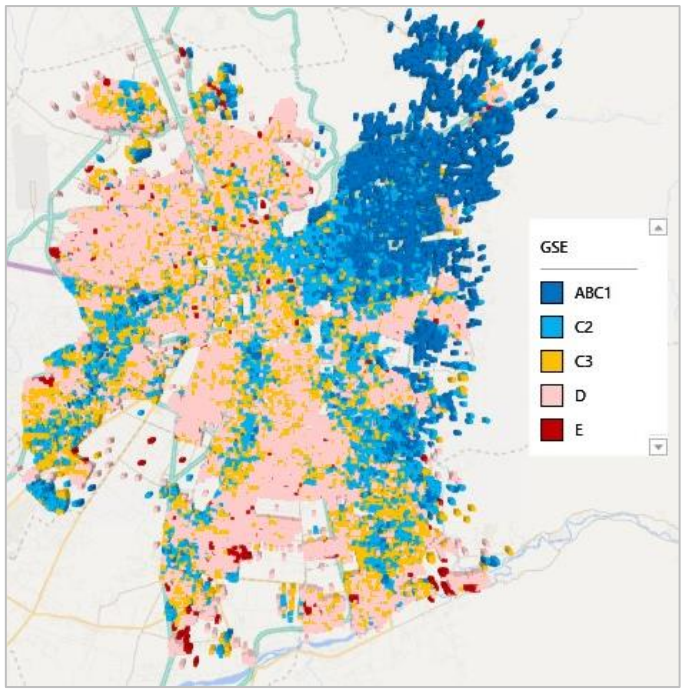
- Tiene **17,6 millones** de habitantes, concentrados principalmente en **10** ciudades
- Tiene una población rural que se estima en un **24%** (OECD), **13%** (oficial)
- Ingreso per cápita: **US$ 22.254**. Índice GINI: **0,508** (Promedio OECD: 0,32)
- Pueblos originarios: **11,08%** de la población (Censo 2012)

- Territorio largo y angosto: 4.329 Km de largo (39 grados de latitud), 180 km ancho (promedio). Superficie continental: 756.770 km^2
- Clima: 7 subtipos climáticos
- Altamente sísmico

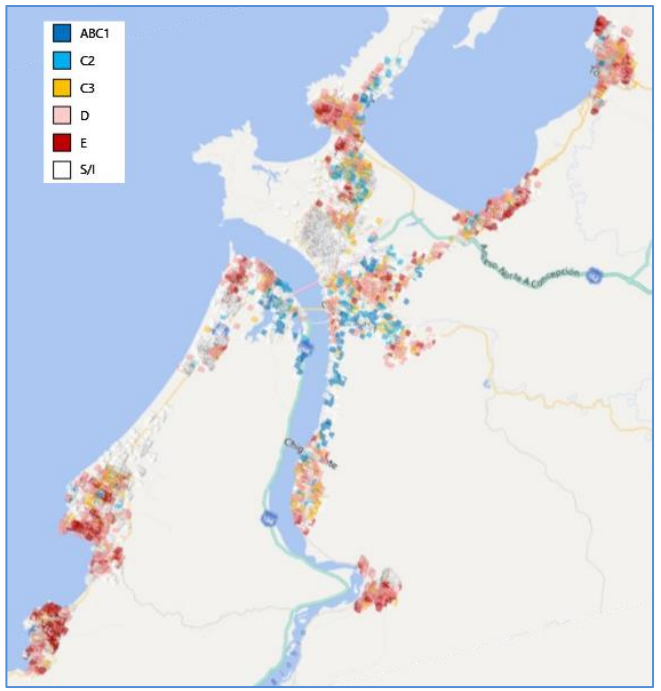
© 2015 — Todos los derechos reservados

1 Desafíos

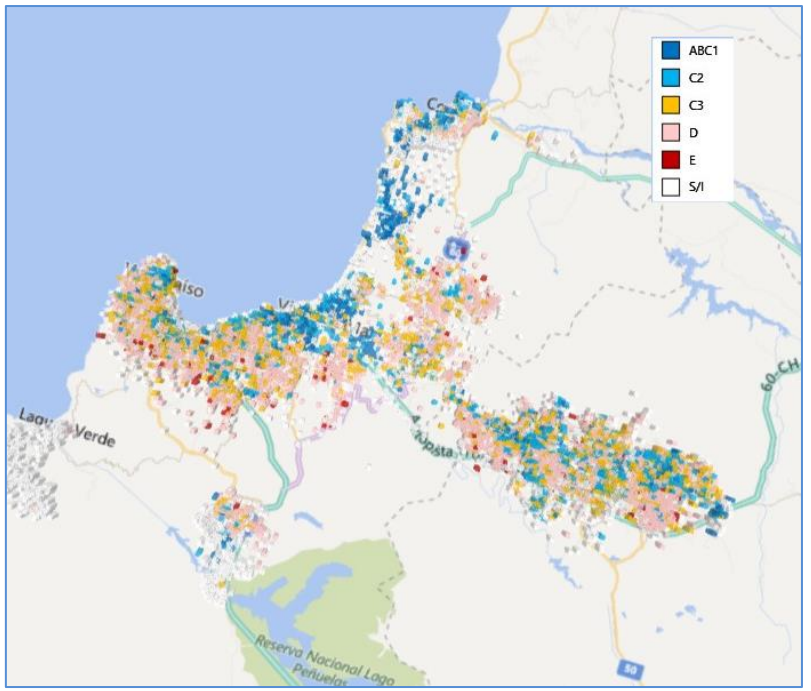
Segregación económica y espacial



Santiago
5.896.836 hab.



Gran Concepción
965.199 hab.



Gran Valparaíso
936.127 hab.

© 2015 — Todos los derechos reservados

1

Desafíos

Retos para la nueva institucionalidad

La diversidad de Chile genera retos para los nuevos Servicios Locales de Educación:

- Diferencias geográficas y socio-demográficas
- Accesibilidad urbana
- Concentración de los centros de empleo
- Segregación socioeconómica, entre otros

Esto afecta al sistema educativo, porque opera en un contexto de relaciones funcionales entre territorios

Sin embargo, estas relaciones no son evidentes y existe el riesgo se ignoren en la toma de decisiones

¿Cómo podemos visibilizar esas relaciones para integrarlas a las políticas públicas?

© 2015 — Todos los derechos reservados

2 Oportunidades

© 2015 — Todos los derechos reservados

2 Oportunidades

Data science

¿Qué? : área multidisciplinaria que combina los campos de Ciencia de la Computación, Matemáticas y Estadística, Econometría e Investigación Operativa y el conocimiento específico de un dominio

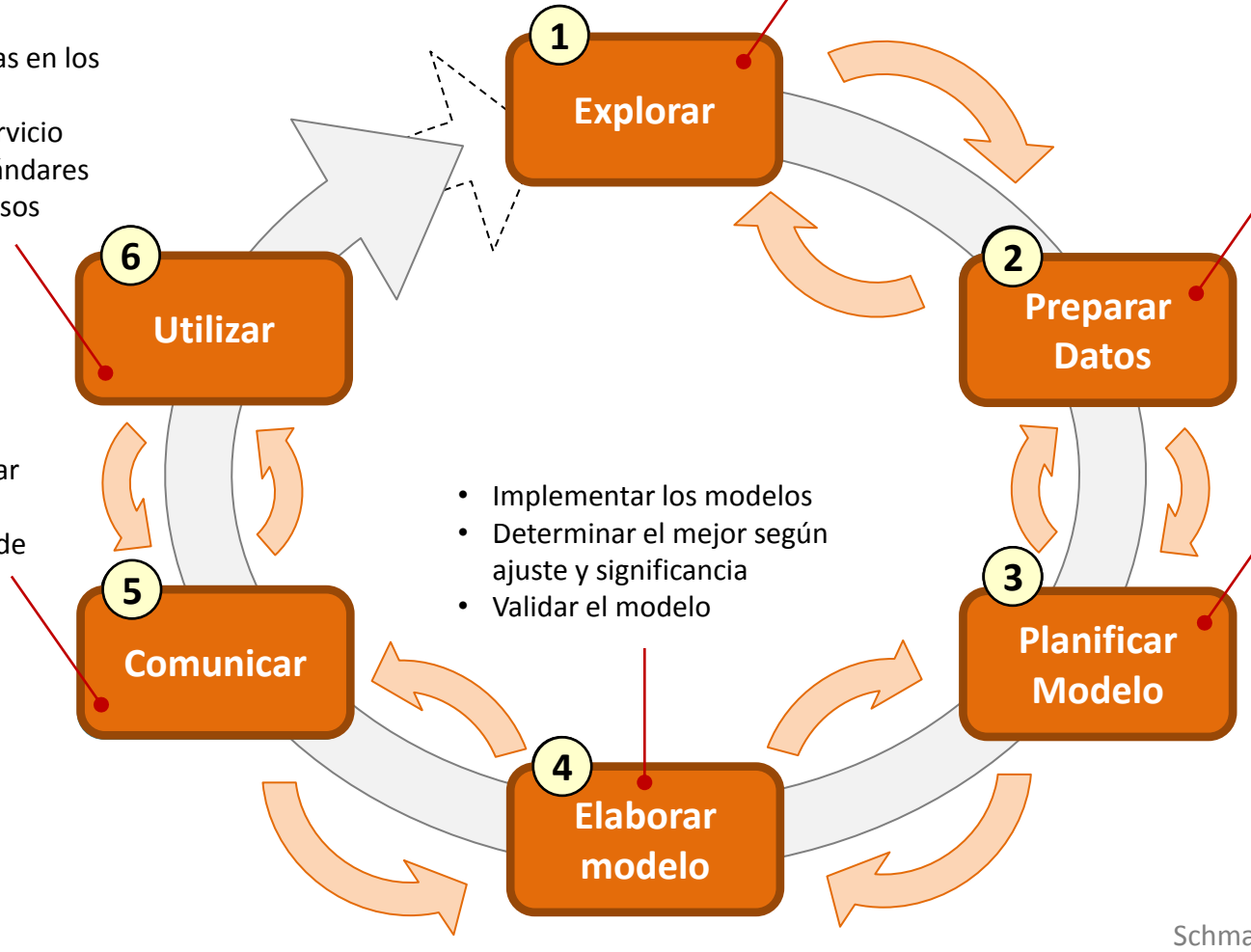
¿Para qué? : para realizar análisis avanzados para informar la toma de decisiones

¿Cómo? : combinando distintas fuentes de datos, para construir modelos que expliquen o predigan ciertos fenómenos o resultados

© 2015 — Todos los derechos reservados

2 Oportunidades

Ciclo de vida del análisis de datos



- Explorar los datos disponibles
- Formular la pregunta
- Formular hipótesis

- Determinar datos necesarios
- Recopilar datos
- Limpiar datos
- Analizar consistencia de datos

- Determinar qué variables explican y cuáles predecir
- Seleccionar posibles modelos y algoritmos a utilizar
- Definir métricas de desempeño

- Implementar los modelos
- Determinar el mejor según ajuste y significancia
- Validar el modelo

- Interpretar resultados
- Generar visualizaciones adecuadas para comunicar resultados
- Hacer recomendaciones de mejora

- Tomar decisiones basadas en los resultados
- Definir estándares de servicio
- Metas para alcanzar estándares
- Planificar y asignar recursos

Schmarzo (2012): Data Analytics Lifecycle

© 2015 — Todos los derechos reservados

2 Oportunidades

Explorar: Desplazamiento diario de estudiantes

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	AGNO	RBD	DGV_RBD	NOM_RBD	COD_REG_RI	COD_PRO_R	COD_COM_F	NOM_COM_	COD_DEPE	COD_DEPE2	RURAL_RBD	COD_ENSE	COD_ENSE2	COD_ENSE3	COD_GRADC	COD_GRADC	LET_CUR	COD_JC
2	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
3	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
4	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
5	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
6	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
7	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
8	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
9	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
10	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
11	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	3	3	A	
12	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
13	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
14	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
15	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
16	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
17	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
18	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
19	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
20	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	463	8	7	4	4	A	
21	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	510	7	6	1	1	A	
22	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	510	7	6	1	1	A	
23	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	510	7	6	1	1	A	
24	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	510	7	6	1	1	A	
25	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	510	7	6	1	1	A	
26	2013	1	9	LICEO POLITE	15	151	15101	ARICA	2	1	0	510	7	6	1	1	A	

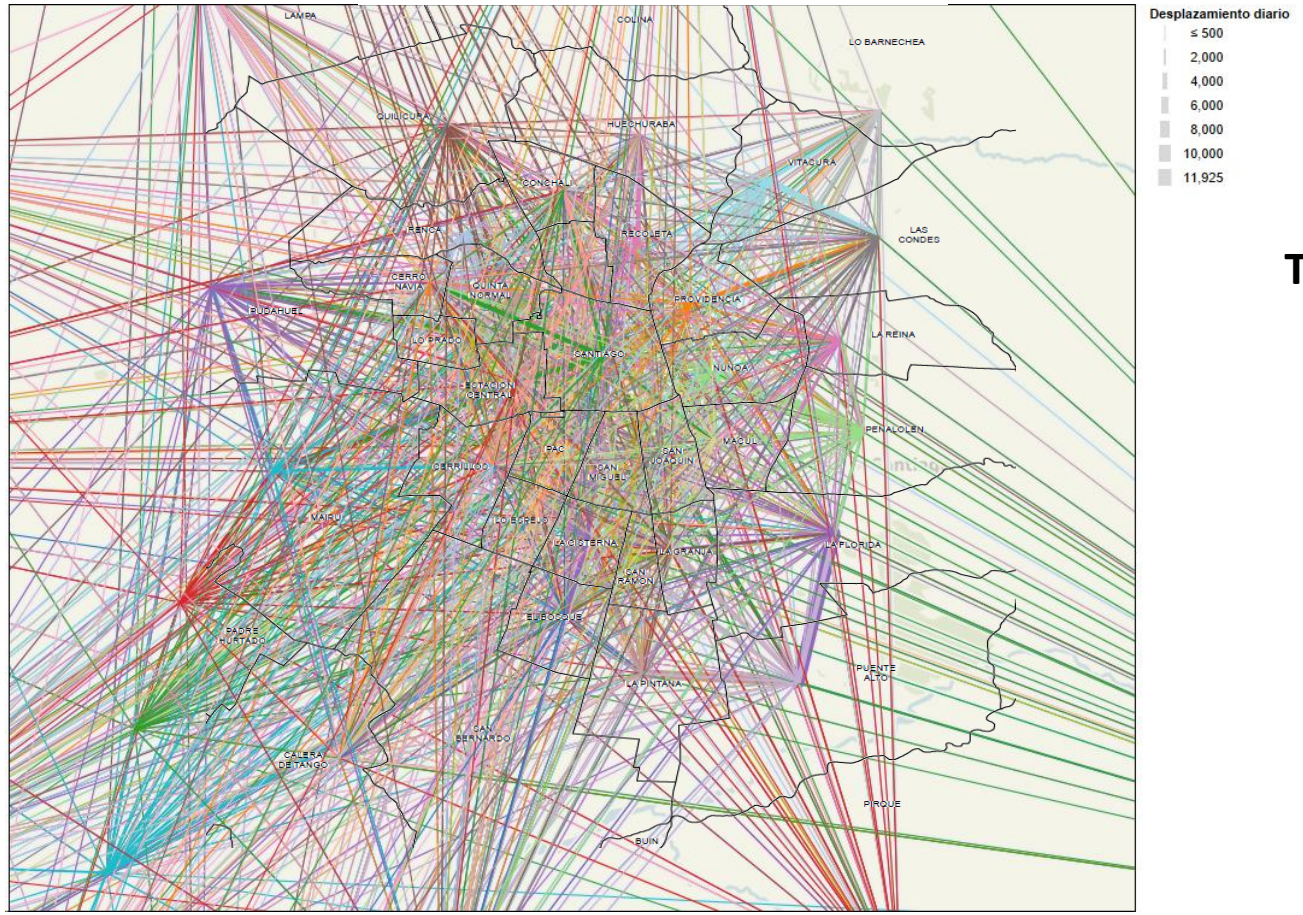
20130917_Matrícula_única_2013_2

Dataset matrícula única 2013,
Centro de Estudios Mineduc

© 2015 — Todos los derechos reservados

2 Oportunidades

Explorar: Desplazamiento diario de estudiantes



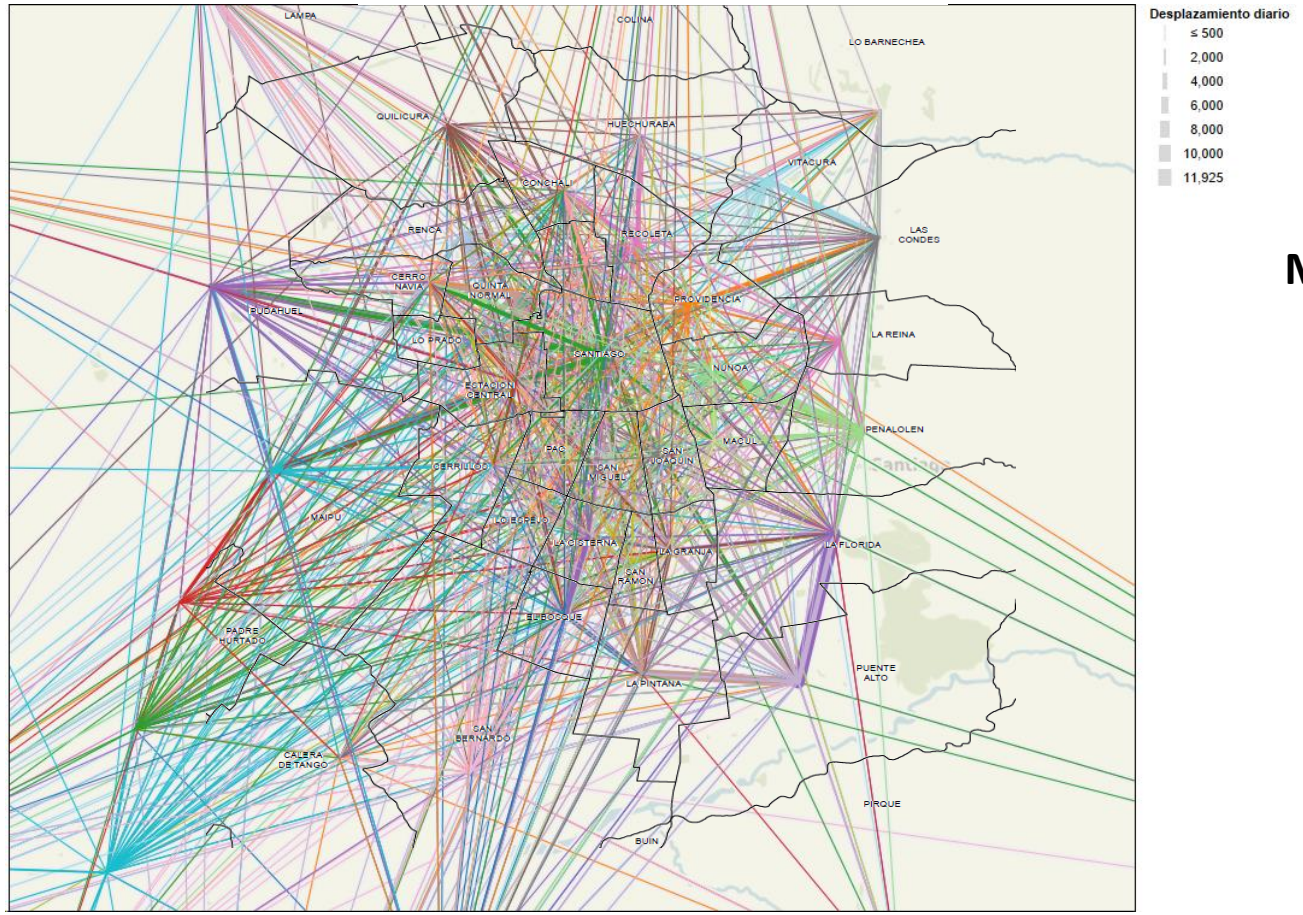
Todos

Dataset matrícula única 2013,
Centro de Estudios Mineduc

© 2015 — Todos los derechos reservados

Oportunidades

Explorar: Desplazamiento diario de estudiantes



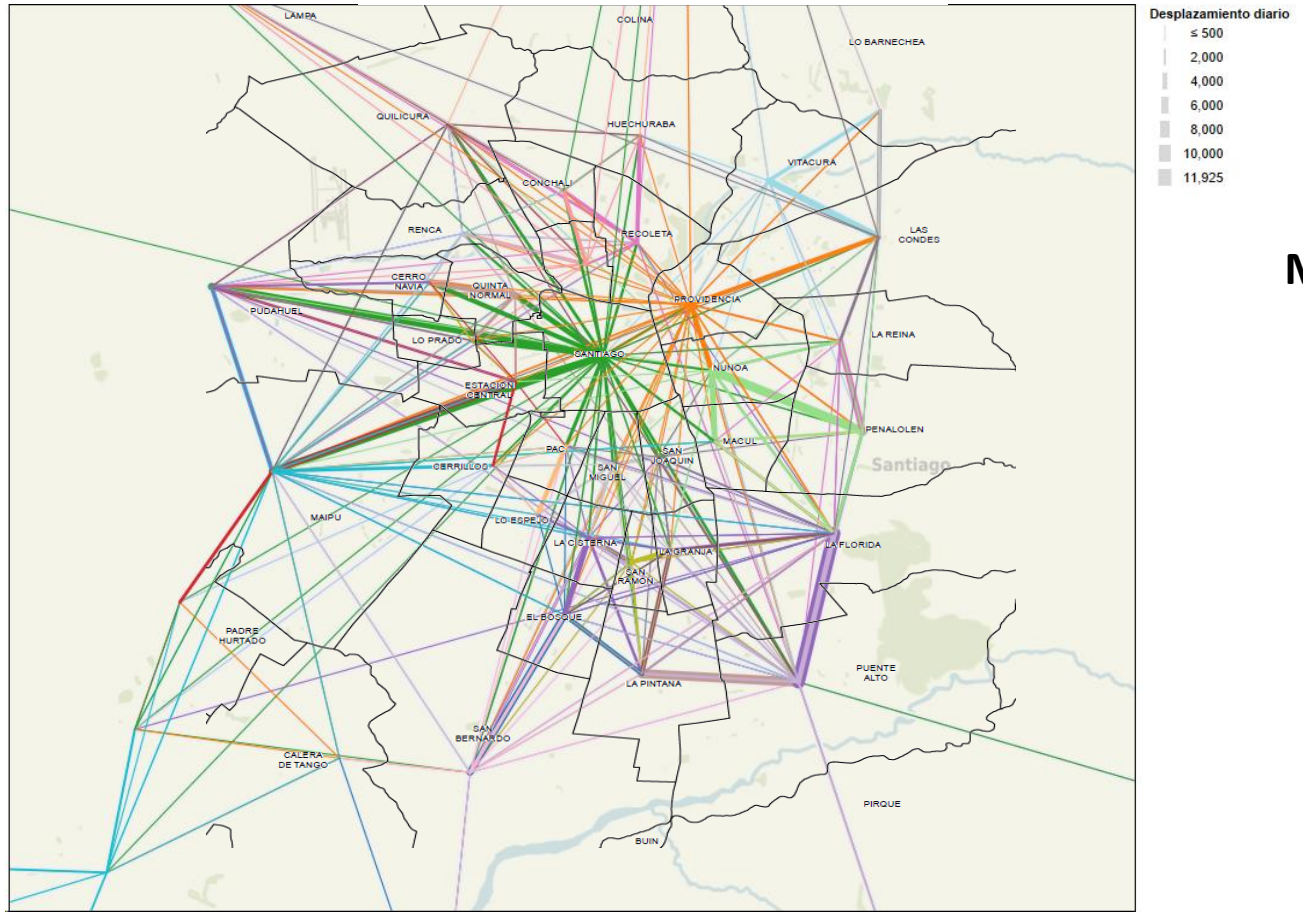
Mayores que 10

Dataset matrícula única 2013,
Centro de Estudios Mineduc

© 2015 — Todos los derechos reservados

2 Oportunidades

Explorar: Desplazamiento diario de estudiantes



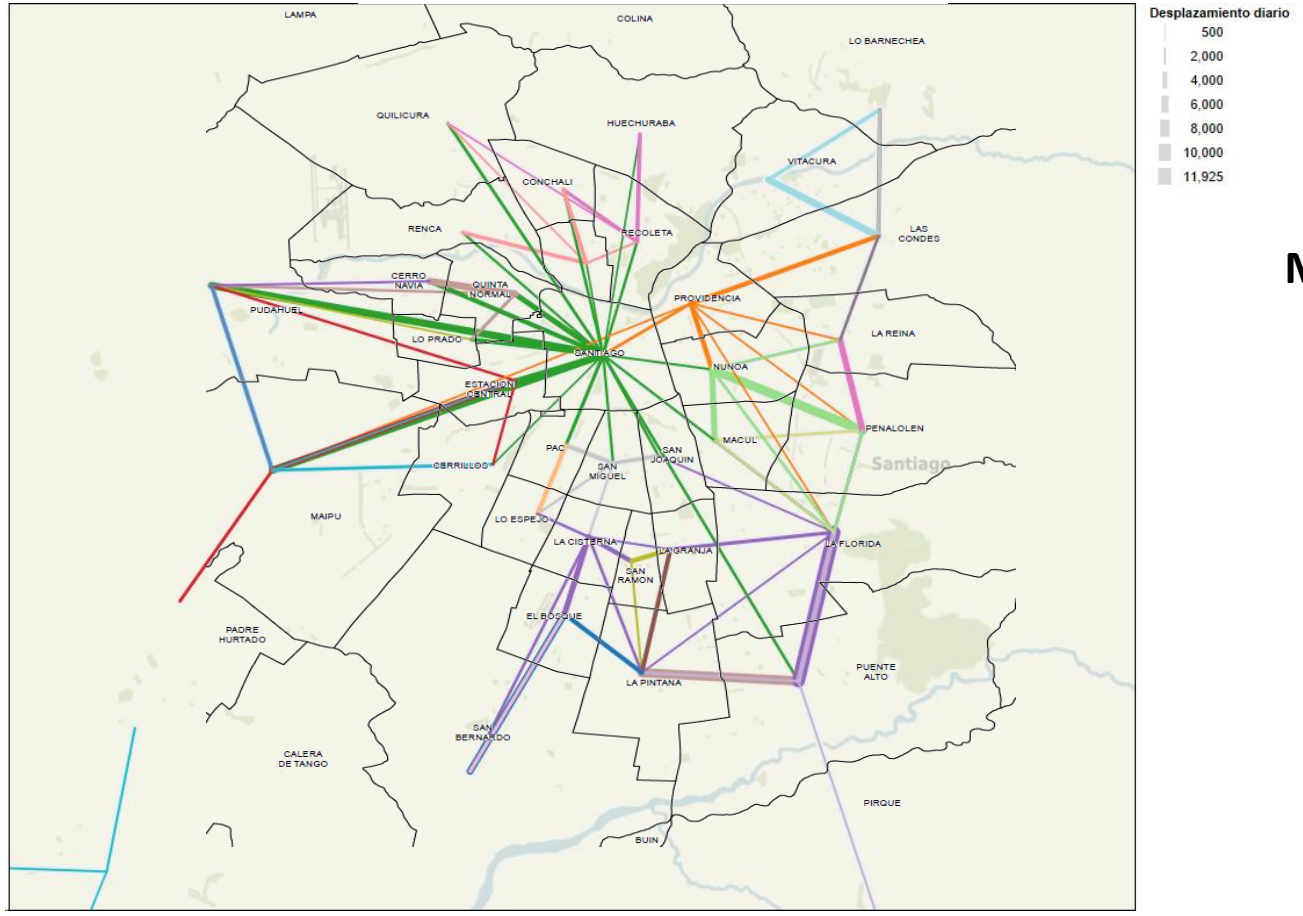
Mayores que 100

Dataset matrícula única 2013,
Centro de Estudios Mineduc

© 2015 — Todos los derechos reservados

2 Oportunidades

Explorar: Desplazamiento diario de estudiantes



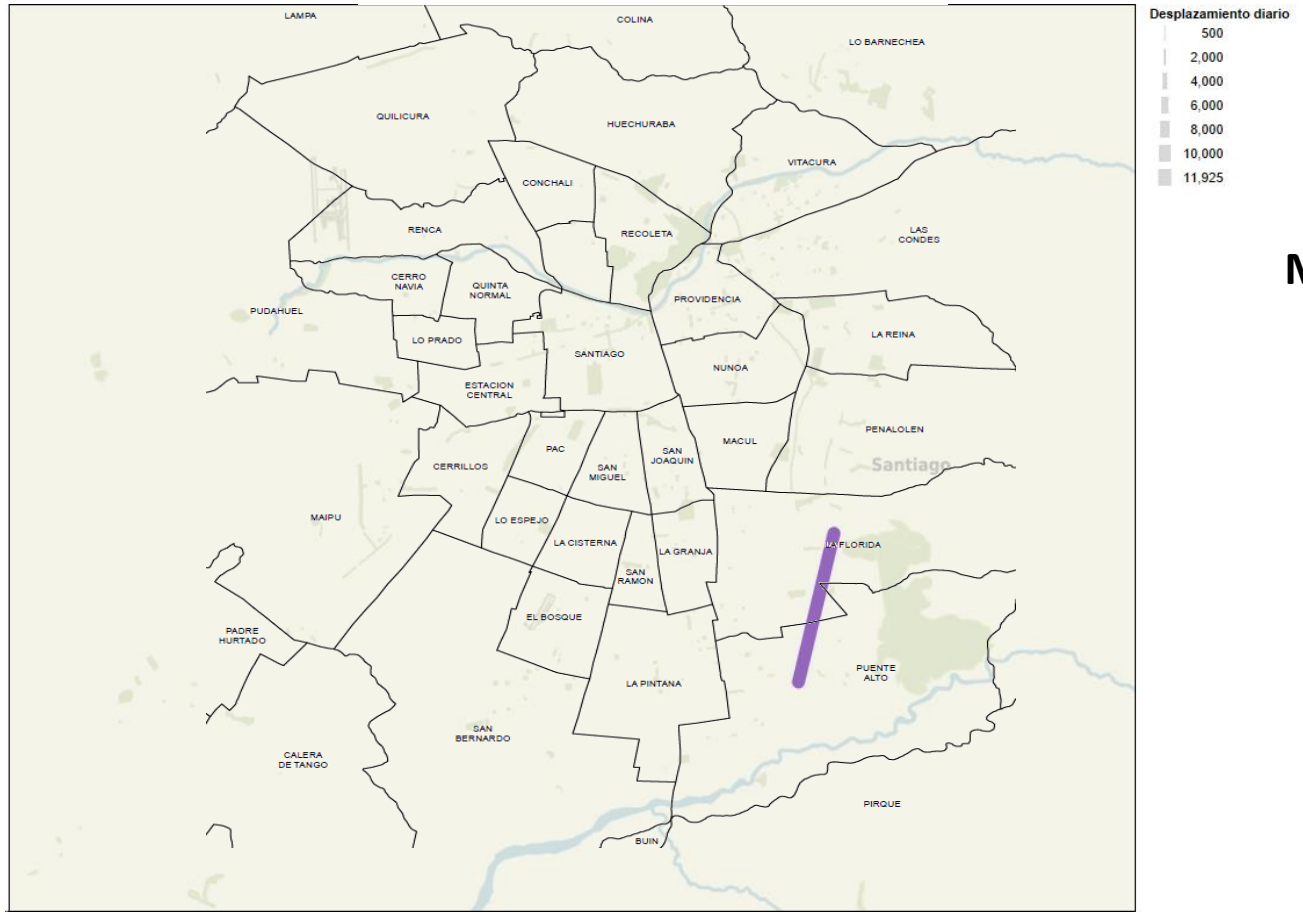
Mayores que 1.000

Dataset matrícula única 2013,
Centro de Estudios Mineduc

© 2015 — Todos los derechos reservados

2 Oportunidades

Explorar: Desplazamiento diario de estudiantes



Mayores que 10.000

Dataset matrícula única 2013,
Centro de Estudios Mineduc

© 2015 — Todos los derechos reservados

2

Oportunidades

Formulación de la pregunta

¿Por qué algunos estudiantes estudian fuera de su comuna?

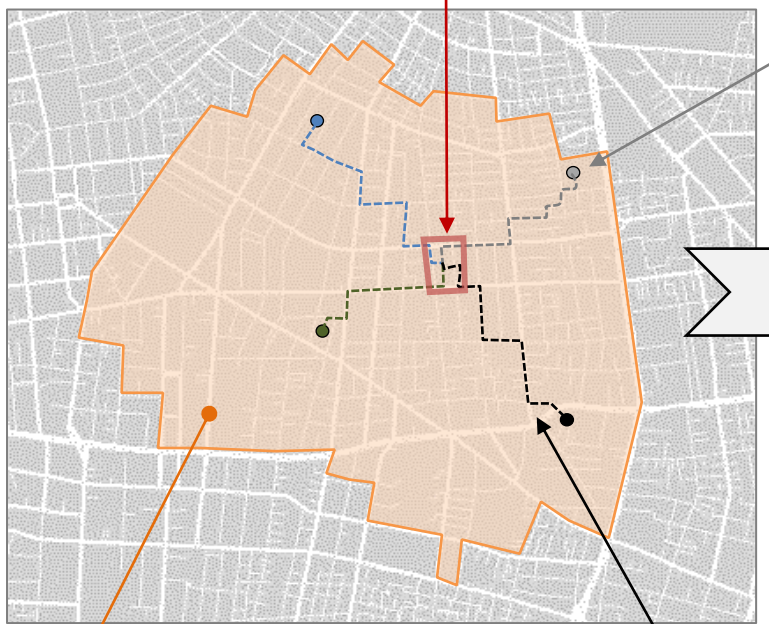
Para ello examinaremos cuál es la cobertura actual de matrícula en función de la situación actual y accesibilidad inmediata

© 2015 — Todos los derechos reservados

2 Oportunidades

Planificar modelo

Nodo demanda
(Ej: alumnos por manzana)



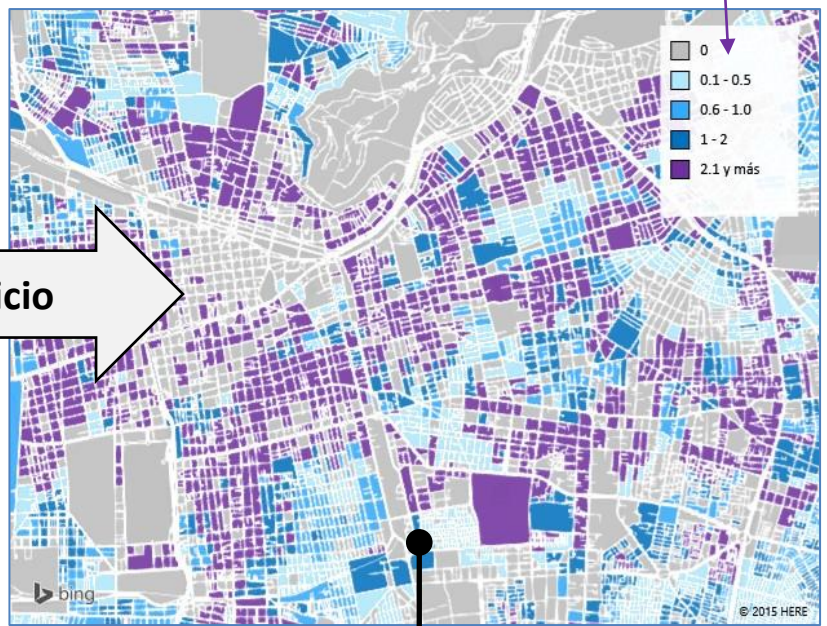
Nodo oferta
(Ej: colegio)

$$E_k = \frac{vacantes_k}{\sum_{j \in I_k} Pm_j}$$

Indicador de servicio

$$ISE_{m_j} = \sum_{k \in I_{m_j}} E_k$$

Colores indican nivel de servicio a nivel de manzana (puntos de corte)



Isocrona:
Lugar geométrico de todos los lugares que se pueden alcanzar desde un punto dado, en un **tiempo fijo**, en un medio de transporte (ej: **auto**)

Trayectoria:
Nodo oferta a nodo demanda según modelo de transporte

Produce **Indicadores de contexto** dónde estudian o viven estudiantes que son **entradas** para **modelos explicativos y/o predictivos** usando análisis econométricos y aprendizaje automático

© 2015 — Todos los derechos reservados

2 Oportunidades

Elaborar modelo

Índice de Servicio Escolar (ISE)

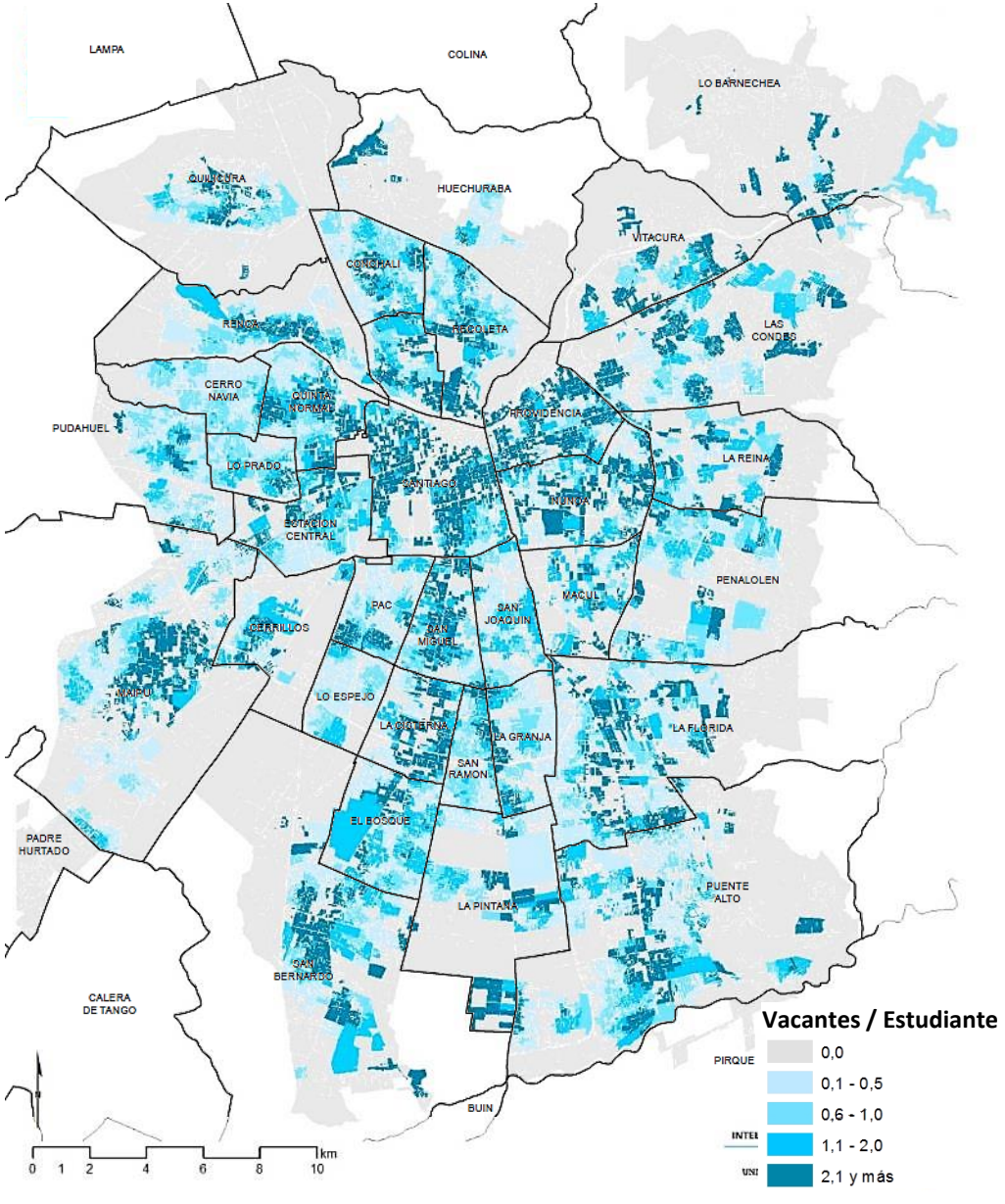
La tasa de **oferta** (vacantes) vs **demanda** (población escolar) en una isocrona de a **10 minutos** caminando

Datos:

- Matrícula 2012
- Censo 2002

Hay **412.819** (**32%**) sin acceso a establecimientos educacionales

¡Por eso la gente se desplaza!



© 2015 — Todos los derechos reservados

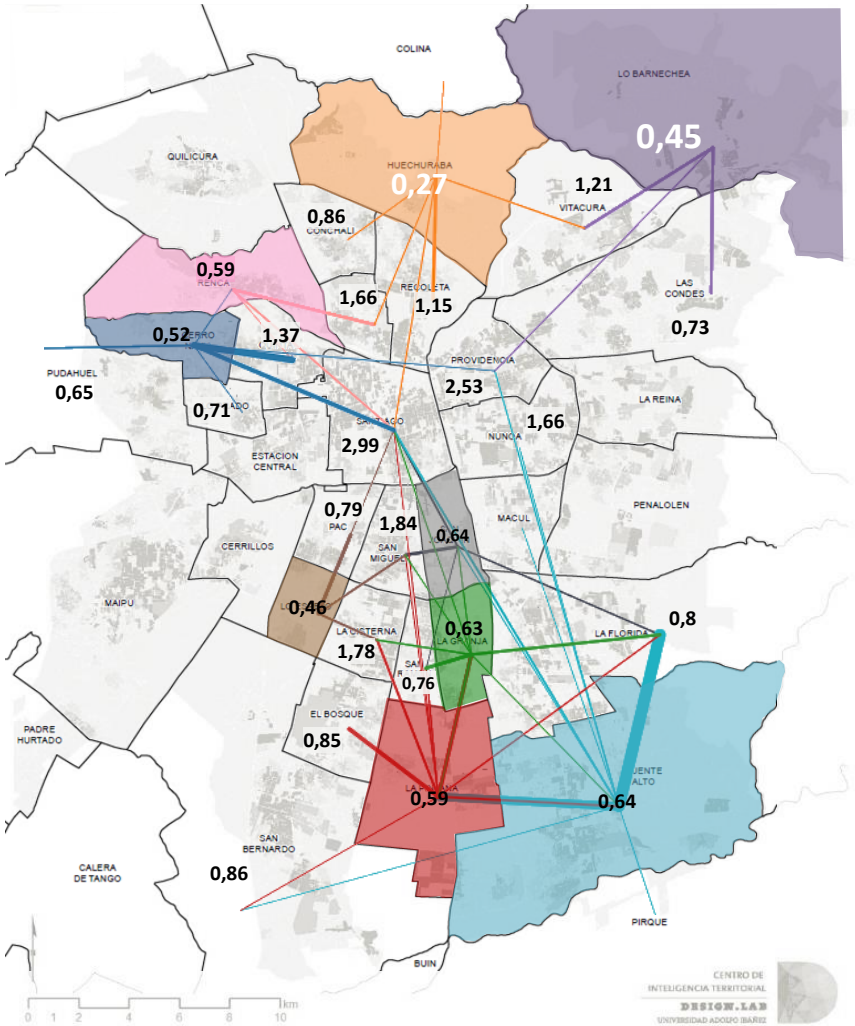
2 Oportunidades

Comunicar resultados

¿Dónde estudian los alumnos con las peores coberturas?

Mejores		Peores	
Santiago	2,99	Huechuraba	0,27
Providencia	2,53	Lo Barnechea	0,45
San Miguel	1,84	Lo Espejo	0,46
La Cisterna	1,78	Cerro Navia	0,52
Ñuñoa	1,66	La Pintana/Renca	0,59
Independencia	1,66	La Granja	0,63
Vitacura	1,21	San Joaquín	0,64
Recoleta	1,15	Puente Alto	0,64

Nivel ciudad: **0,82**



© 2015 — Todos los derechos reservados

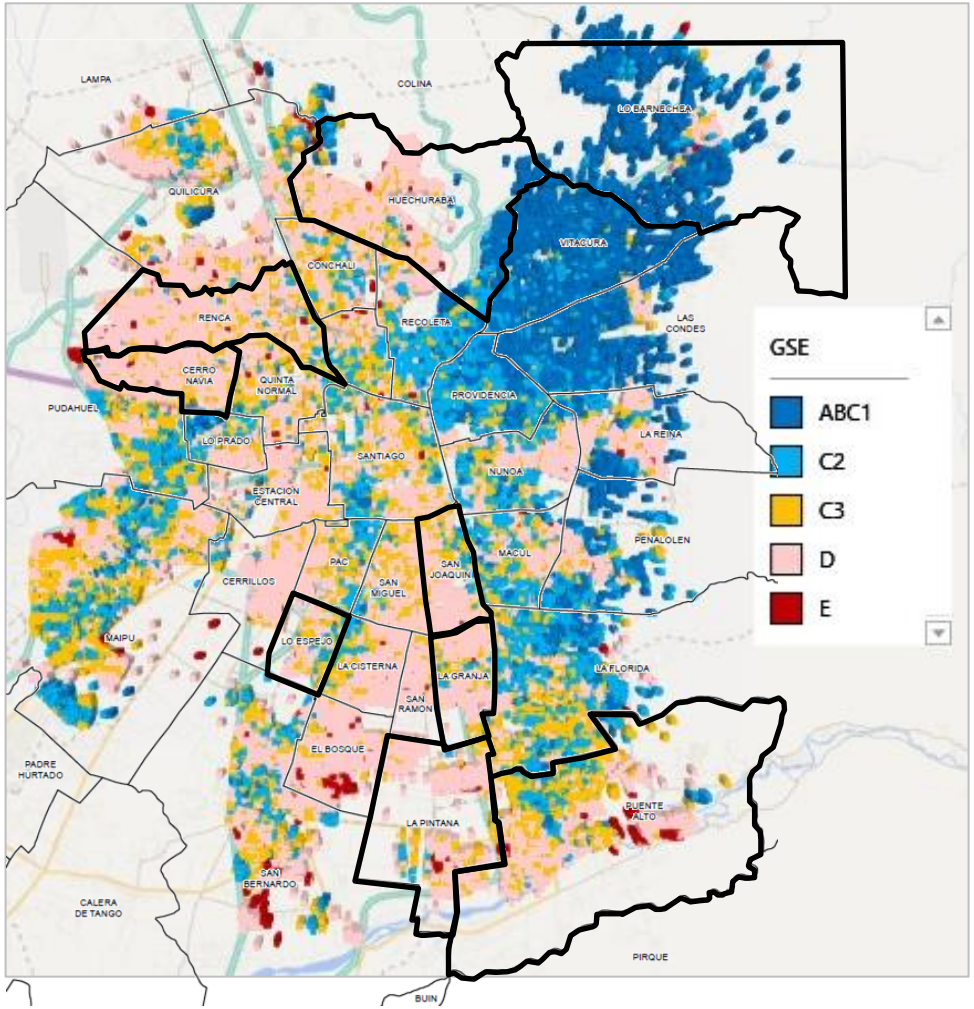
2 Oportunidades

Comunicar resultados

¿Dónde estudian los alumnos con las peores coberturas?

Mejores		Peores	
Santiago	2,99	Huechuraba	0,27
Providencia	2,53	Lo Barnechea	0,45
San Miguel	1,84	Lo Espejo	0,46
La Cisterna	1,78	Cerro Navia	0,52
Ñuñoa	1,66	La Pintana/Renca	0,59
Independencia	1,66	La Granja	0,63
Vitacura	1,21	San Joaquín	0,64
Recoleta	1,15	Puente Alto	0,64

Nivel ciudad: **0,82**



© 2015 — Todos los derechos reservados

2 Oportunidades

Síntesis

La **dimensión territorial** da cuenta de la **heterogeneidad** del país y fundamental en la reforma educativa

Usando *Data Science* podemos explicitar fenómenos **invisibles** para la implementación práctica de la reforma, en términos de:

- **Diagnosticar brechas**
- Definir **estándares de servicio**

Podemos establecer **prioridades**, asignar **recursos** y definir **metas** para disminuir las **brechas** que puedan existir en los territorios de un SLE

© 2015 — Todos los derechos reservados

2 Oportunidades

Síntesis

Permite **gestionar la transición** hacia el nuevo sistema de administración de la educación pública evitar efectos indeseados:

Ej: la fuga de alumnos del sector público al particular subvencionado agravando el actual déficit estructural de la educación municipal

Es vital contar con información que permita **predecir** tanto **dónde** como **cuánta** será la magnitud de este fenómeno

¿Pero cómo entregamos esta evidencia a los tomadores de decisiones?

© 2015 — Todos los derechos reservados

3 Propuesta

© 2015 —– Todos los derechos reservados

3

Propuesta

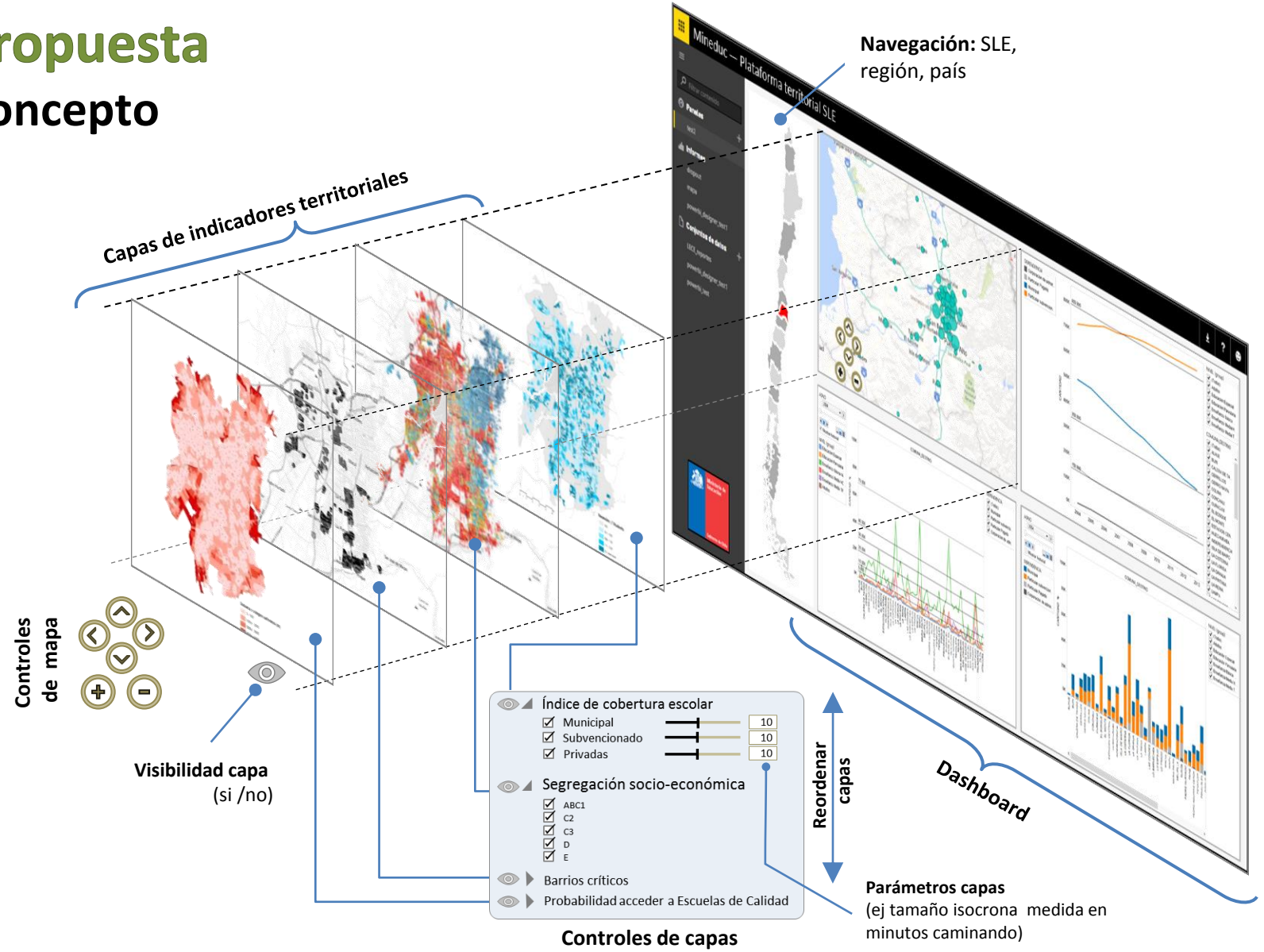
Cómo entregar la evidencia

A través de una **plataforma en línea** que apoye la gestión de los nuevos Servicios Locales de educación

- Otorgando **información** oportuna, objetiva, y actualizada
- Que genere predicciones sobre ciertos fenómenos de interés
- Que apoye la correcta **toma de decisiones** respecto de la implementación de esta nueva política pública en educación

© 2015 — Todos los derechos reservados

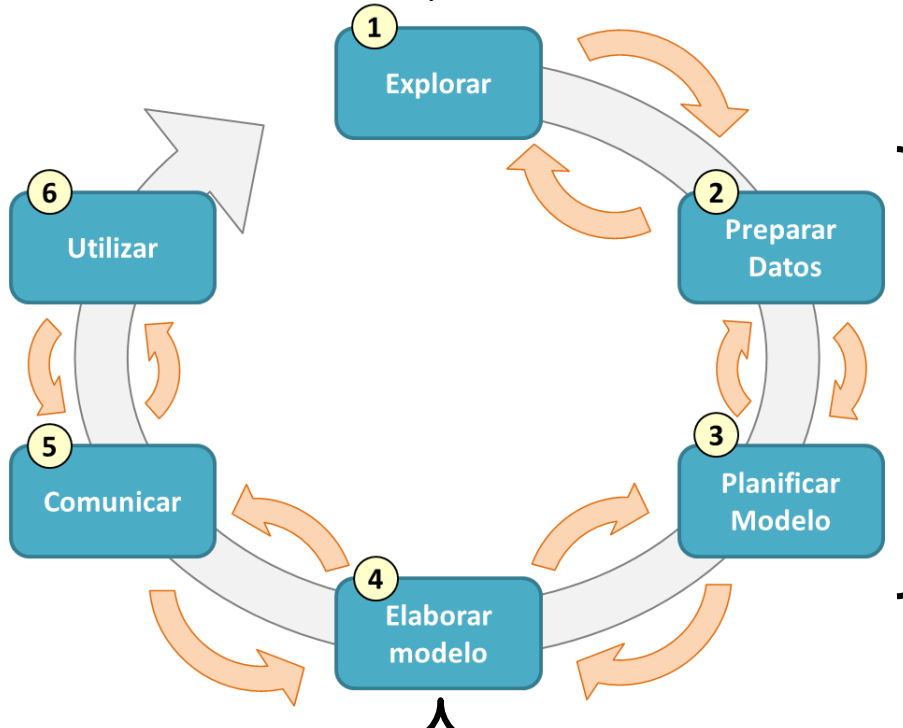
3 Propuesta
Concepto



© 2015 — Todos los derechos reservados

3 Propuesta
Tecnología

- Power BI Desktop / Power BI web
- Excel (3D maps)
- SQL *Server*
- **QGIS (open source)**



- Power BI Desktop / Power BI web
- Excel (3D maps)
- SQL *Server*
- **QGIS**

- Excel
- Azure Cloud Services:
  - Azure Virtual Machines
  - SQL Databases
  - Data Factory
  - Azure Storage Accounts

- Azure Machine Learning Studio
- **R**
- Excel

Schmarzo (2012): Data Analytics Lifecycle

© 2015 — Todos los derechos reservados

3 Propuesta

Tecnología — Azure ML Studio

Ejemplo: predecir dónde los estudiantes de **primaria** que deben cambiarse de colegio cuando la oferta del siguiente grado no existe (Santiago)

Opciones de cambio: de un colegio público a uno particular-subvencionado o viceversa

Variables:

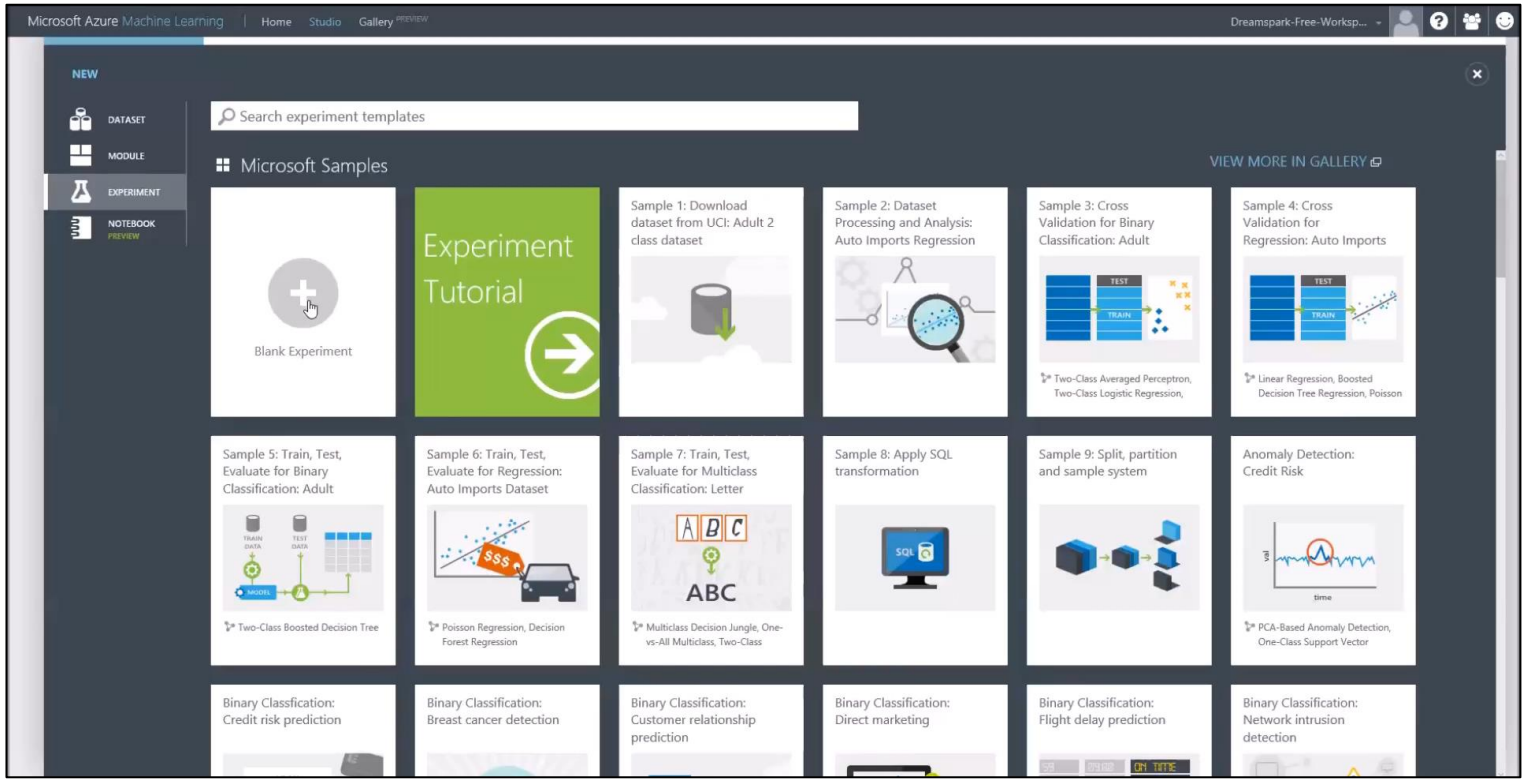
- Si se cambio de domicilio
- Notas (promedio anual)
- Repitencia (si/no)
- Género
- Embarazo
- Edad
- Asistencia anual (%)
- Beneficiario subvención preferencial (si/no)
- Alumno nuevo en el establecimiento actual (si/no)
- Tiene necesidades educativas especiales (si/no)
- Distancia de su domicilio al establecimiento actual

Cantidad de casos: 17.021 estudiantes (después de preparar los datos)

© 2015 — Todos los derechos reservados

3 Propuesta

Tecnología — Azure ML Studio

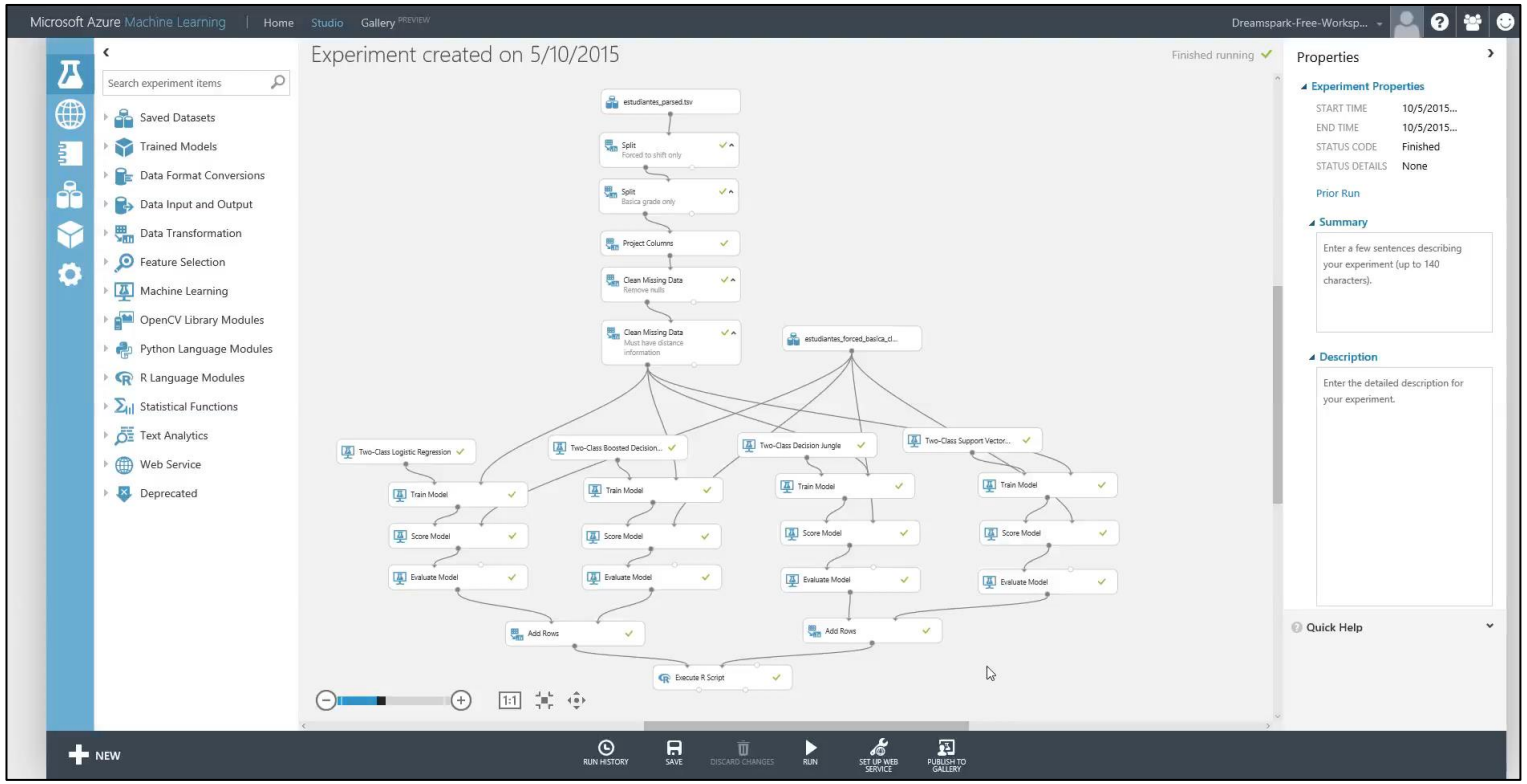


Elaborar modelo — Implementar modelo

© 2015 — Todos los derechos reservados

3 Propuesta

Tecnología — Azure ML Studio



Elaborar modelo — Determinar ajuste

© 2015 — Todos los derechos reservados

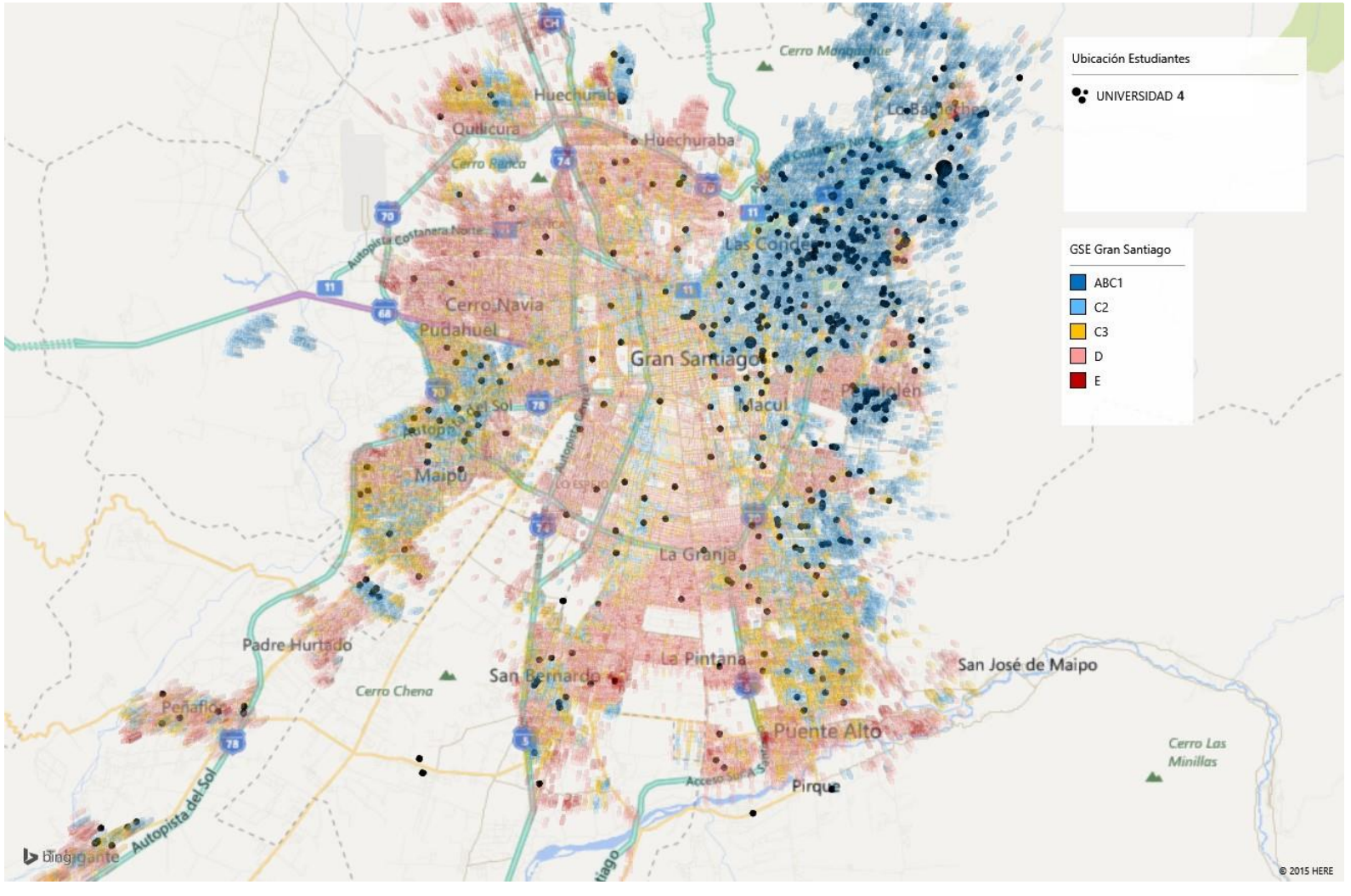
4 Visión

Hace visible lo que antes era invisible por lo tanto, no se podía tomar medidas ('actionable'): aumentar el **uso práctico de la evidencia** para el diseño y evaluación de nuevas políticas públicas

Introduce otros componentes antes ignorados en la toma de decisiones en educación y la necesaria integración a nivel de políticas públicas con otras agencias y Ministerios: **es una herramienta para generar diálogos inter-agencias**

Hacia una **inteligencia de valor público**, que tome decisiones basada en la evidencia y datos que el mismo Estado genera

© 2015 — Todos los derechos reservados



© Francisco Meneses 2012

http://mideuniversidad.cl/

© 2015 — Todos los derechos reservados

Apoyando el diseño y desarrollo de nuevas políticas públicas en educación usando Data Science

Dr. Patricio Rodríguez

prodriguez@ciae.uchile.cl

RedClara

Webminar "El potencial de la analítica avanzada en las Universidades"

Diciembre 02, 2015

Centro de Investigación Avanzada en Educación
Universidad de Chile

UNIVERSIDAD ADOLFO IBÁÑEZ

CONICYT Ministerio de Educación
Gobierno de Chile

FONDEF
Fondo de Fomento al Desarrollo Científico y Tecnológico

Red CLARA
+ Red + Ciencia